Contributions au recalage d'images et à la reconstruction 3D de scènes rigides et déformables

Adrien Bartoli

▶ **To cite this version:**

Adrien Bartoli. Contributions au recalage d'images et à la reconstruction 3D de scènes rigides et déformables. Automatique / Robotique. Université Blaise Pascal - Clermont-Ferrand II, 2008. tel-00344569

HAL Id: tel-00344569

https://theses.hal.science/tel-00344569

Submitted on 5 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ BLAISE PASCAL – CLERMONT II

Ecole Doctorale Sciences Pour l'Ingénieur

HABILITATION À DIRIGER DES RECHERCHES

Préparée au LASMEA, UMR 6602 CNRS / Université Blaise Pascal
(Laboratoire des Sciences des Matériaux pour l'Electronique, et d'Automatique)

Spécialité : Vision par Ordinateur

Présentée et soutenue publiquement par

Adrien Bartoli

le 9 juin 2008

Contributions au recalage d'images et à la reconstruction 3D de scènes rigides et déformables

—— Synthèse des travaux et activités scientifiques sur 2004 – 2007 ——

Contributions to Image Registration and to the 3D Reconstruction of Rigid and Deformable Scenes

—— Scientific Work and Activity over 2004 – 2007 ——

Devant le jury composé de

Président :	Michel Dhome	CNRS (Centre National de la Recherche Scientifique) – LASMEA
Rapporteurs externes :	Pascal Fua	EPFL (Ecole Polytechnique Fédérale de Lausanne)
	Richard Hartley	ANU (the Australian National University)
	Nikos Paragios	ECP (Ecole Centrale de Paris)
Rapporteur interne :	Jean-Marc Lavest	UdA (Université d'Auvergne) – LASMEA
Examinateurs :	Mads Nielsen	DIKU (Datalogisk Institut, Københavns Universitet)
	Marc Pollefeys	ETHZ (Eldgenössische Technische Hochschule Zürich) et UNC
	Jean Ponce	ENS (Ecole Normale Supérieure) et UIUC

Avant-propos

Ce document comprend deux parties, la première rédigée en français et la deuxième en anglais. Alors que la première partie est une synthèse de l'ensemble de mes activités professionnelles (animation de la communauté scientifique, encadrement, enseignement, projets de recherche, participation à des comités de programme, expertises, publications, *etc.*), la deuxième partie porte sur les aspects purement scientifiques. Un résumé détaillé de la deuxième partie est cependant donné dans la première, ainsi qu'un ensemble de perspectives. La deuxième partie va de pair avec un recueil de mes articles les plus représentatifs dont la structure est similaire. L'annexe A explicite les acronymes des Universités, instituts et groupes de recherche utilisés les plus fréquemment dans ce document.

Ce document couvre la période 2004 – 2007, alors que ma thèse de doctorat couvre en détail mes contributions scientifiques sur la période 2000 – 2003. Les contributions que je décris sont issues de travaux menés à différents endroits : au LASMEA en tant que Chargé de Recherche CNRS, à l'Université d'Oxford avec Andrew Zisserman en tant que post-doctorant, et au DIKU à Copenhague en tant que "Visiting Professor". J'ai travaillé seul sur certains thèmes et au travers de nombreuses collaborations sur d'autres, avec mes étudiants, des étudiants en visite, des stagiaires, et bien sûr d'autres chercheurs.

Il est dans ce cadre important de bien manier la première et la troisième personne. J'exprime, notamment dans l'introduction et les perspectives, des points de vue qui n'engagent que moi. Ils sont parfois basés sur des travaux effectués en collaboration avec d'autres chercheurs, ce qui rend difficile le choix du pronom. J'ai essayé d'indiquer de manière rigoureuse les contributions de chacun.

Forewords

This document comprises two major parts, with the first written in French and the second written in English. The first part summarises all my professional activities, such as my ongoing involvement in the scientific community, supervising, teaching, research projects, participation to program committees, specific expertises and publications. My future perspectives are also described in this first part. The second part is dedicated to outlining the particular scientific aspects of my work. Supplementing this is an accompanying collection of scientific publications which follows the same organization. For clarity, in Appendix A the acronyms for the Universities, research institutes and groups that are the most frequently used in this thesis are defined.

The scope of this document covers my activities over the period of 2004 – 2007, while my PhD thesis provides a detailed description of my scientific contributions over the period 2000 – 2003. My contributions made at various research centres are described, including LASMEA as a CNRS Research Scientist, the University of Oxford as a Post Doctoral fellow under Andrew Zisserman, and DIKU in Copenhagen as a Visiting Professor. In some instances I have worked primarily on my own, whilst in others there has been considerable collaboration with other researchers. These have included my students, visiting students, trainees and other researchers.

Throughout this document I have attempted to make clear the distinction between my own work and the collaborative work done with other researchers. In particular in the Introduction and Perpectives Sections my own thoughts are presented which may not necessarily be shared by other researchers involved with my work.

Remerciements – Acknowledgments

… and many others!

Table des matières

Première partie

Activités scientifiques et administratives, et synthèse des travaux de recherche

CHAPITRE

1

ACTIVITÉS SCIENTIFIQUES ET ADMINISTRATIVES

Ce chapitre présente une synthèse de mes activités scientifiques et d'administration de la recherche. Il couvre la présentation de l'équipe de recherche ComSee ("Computers that See") dont j'ai la co-responsabilité, mes activités d'encadrement et d'enseignement. Les autres aspects abordés sont mes activités d'animation de la communauté scientifique par l'organisation de congrès, ateliers et tutoriels, mes responsabilités diverses, et se termine par la liste de mes publications.

1.1 Récapitulatif

Je donne ci-dessous un récapitulatif de mes contributions et activités.

Publications

Les revues ne sont pas classées par date de publication effective mais plutôt par période de réalisation des travaux, même si certains de mes travaux de thèse (période 2000 – 2003) n'ont été soumis qu'a posteriori à des revues. Ces chiffres ne prennent pas en compte les 5 articles en cours d'expertise dans des revues.

	Total	2000 – 2003	2004 – 2007
Revues internationales	12	8	4
Articles invités	1	0	1
Congrès et ateliers internationaux	50	16	34
Congrès et ateliers nationaux	15	5	10
Rapports de recherche	4	4	0
Totaux	82	33	49
Premier / deuxième auteur sur	51% / 38%	79% / 18%	34% / 51%

J'ai donné 4 présentations invités dans des congrès ou ateliers, et 22 séminaires invités dans des laboratoires lors de visites.

Encadrements

J'indique les étudiants que j'ai encadrés depuis 2004 ou encadre actuellement. Je suis par ailleurs actuellement co-encadrant d'un post-doc et rapporteur pour 2 thèses.

	Plus de 95%	Entre 30% et 95%	Moins de 30%
Niveau thèse	3	4	1
Niveau Master 2 Recherche	2	4	1
Niveau inférieur	5	2	2

Enseignements

Je donne ci-dessous le nombre d'heures de cours et de TDe (Travaux Dirigés d'expérimentation) que j'ai dispensées par année.

	2003 / 2004	2004 / 2005	2005 / 2006	2006 / 2007	2007 / 2008
Cours	0	17	31,5	56,5	66+
TDe	10	0	20	20	0

Projets de recherche

Je participe à 3 projets financés par l'Agence Nationale de la Recherche (ANR). Parmi ceux-ci on trouve 2 projets jeunes chercheurs et 1 projet blanc. Je participe ou ai participé à 5 projets bilatéraux sur 2004 – 2007. J'ai bénéficié d'un échange de chercheurs CNRS en 2007 et je suis actuellement impliqué dans un contrat de recherche avec le CEA.

Expertises et participation à des comités de programme

J'expertise des articles pour de nombreuses revues et congrès. J'ai participé à 4 comités de programme en 2006, 9 en 2007 et 9 pour l'instant en 2008. J'ai co-organisé 2 ateliers, 1 congrès et 1 tutoriel. J'effectue des expertises pour plusieurs Appels à Projets de l'ANR.

Autres éléments

Je suis co-responsable de l'équipe ComSee et "Visiting Professor" au DIKU à Copenhague sur 2006 – 2009. Je siège dans divers commissions au niveau du LASMEA et des Universités Clermontoises. Je fais de la consultance auprès de la société FittingBox de Toulouse depuis 2007.

1.2 Parcours universitaire et professionnel

1.2.1 Détails personnels

Je suis né le 9 avril 1977 à Grenoble. Je suis célibataire et père de deux enfants.

1.2.2 Distinctions

J'ai reçu le prix de thèse de l'Institut National Polytechnique de Grenoble (INPG) en 2004, et le prix du meilleur article au congrès CORESA en 2007.

1.2.3 Parcours universitaire

J'ai fais mon parcours universitaire à Grenoble, à l'Université Joseph Fourier (UJF) et à l'INPG. J'ai obtenu les diplômes suivants :

2003	**Doctorat en vision par ordinateur**, INPG Titre de la thèse : *Reconstruction et alignement en vision 3D : points, droites, plans et caméras*
2000	**DEA en vision, synthèse d'image et robotique**, INPG (major de promotion)
2000	**Magistère d'informatique**, UJF (major de promotion)
1999	**Maîtrise d'informatique**, UJF (major de promotion)
1998	**Licence d'informatique**, UJF (major de promotion)

1.2.4 Parcours professionnel

2006-	**Co-responsable de ComSee**, équipe de 9 permanents du GRAVIR / LASMEA
2004-	**Chargé de Recherche au CNRS** affecté au GRAVIR / LASMEA (reçu premier au concours national)
2006-2009	**"Visiting Professor"** au DIKU, Université de Copenhague
2003-2004	**Chercheur post-doctoral** à l'Université d'Oxford, avec Andrew Zisserman
2000-2003	**Etudiant en thèse** à l'INRIA, équipe Perception, avec Peter Sturm et Radu Horaud
Eté 2000	**Stagiaire** à l'Université d'Oxford, avec Andrew Zisserman
1999-2003	**Moniteur** en informatique à l'UJF

1.3 L'équipe de recherche ComSee / GRAVIR / LASMEA

Je donne ci-dessous un historique récent et la structuration de l'équipe de recherche ComSee dont j'ai la co-responsabilité depuis septembre 2006.

1.3.1 Contexte

Je suis arrivé au LASMEA en octobre 2004 suite à mon séjour post-doctoral au Visual Geometry Group à l'Université d'Oxford. A cette époque, la structuration en thèmes ou équipes de GRAVIR n'avait pas d'incidence hiérarchique, du moins pas explicitement. Il s'est tenu en 2005 une Assemblée Générale durant laquelle l'organisation du groupe a été modifiée : sa direction est désormais assurée par un responsable au lieu de trois précédemment. Il est apparu naturel de structurer le groupe en trois équipes de recherche, autour des thèmes suivants : systèmes de perception, vision par ordinateur et robotique. Jean-Marc Lavest a accepté la responsabilité de l'équipe de vision par ordinateur qu'il a alors rebaptisée, sur une suggestion de ma part, "Computers that See", ou "ComSee". En septembre 2006, Jean-Marc a été élu responsable du groupe GRAVIR. J'ai alors repris la responsabilité de ComSee conjointement avec Thierry Chateau. Nous assurons ce rôle depuis lors, en interaction étroite avec les responsables des autres équipes de recherche, Roland Chapuis puis Jean-Pierre Dérutin pour "PerSyst" (systèmes de perception), Benoît Thuilot puis Philippe Martinet pour "ROSACE" (robotique), et Laurent Trassoudaine, responsable du groupe GRAVIR depuis avril 2007.

1.3.2 Volume d'activité

ComSee compte 9 chercheurs et enseignants-chercheurs permanents "actifs", ce qui correspond à approximativement 4,5 Equivalents Temps Plein, et une vingtaine de doctorants. Les personnes Habilitées à Diriger des Recherche sont au nombre de 3. L'activité de l'équipe a fortement augmentée sur les quelques dernières années. Sur le dernier contrat quadriennal, 2002 – 2006, 8 thèses ont été soutenues ; nous prévoyons environ 20 soutenances pour la période 2006 – 2012. ComSee est actuellement impliquée dans 7 projets de recherche financés par l'ANR et plusieurs contrats de recherche, avec notamment le CEA, Renault et la DGA.

1.3.3 Structuration scientifique

Peu après avoir pris la responsabilité de ComSee, Thierry Chateau et moi-même avons re-structuré l'équipe, afin de refléter l'état des forces vives et des problématiques scientifiques abordées et afin de poser explicitement les éléments de cohérence scientifique. L'équipe est structurée autour des deux problématiques suivantes :

1. **Mise en relation d'images.** Etant données plusieurs images de la même scène (ou plusieurs images présentant un contenu sémantique similaire), comment mettre en correspondance les pixels de ces images, ou des primitives géométriques tels des points d'intérêt qui en sont extraits ?
2. **Reconstruction 3D.** Etant données une ou plusieurs images d'une même scène, comment représenter et estimer la structure 3D observée et la position de la caméra ?

Ces deux problématiques sont déclinées autour de trois axes de recherche principaux :

1. **Reconstruction 3D de scènes rigides et métrologie par vision** (responsable : Maxime Lhuillier, Chargé de Recherche CNRS). Le but de cet axe est l'étude de méthodes pour la vision dans un environnement supposé statique, avec des applications telles que la localisation et la cartographie, éventuellement réalisées simultanément. Il est important que les méthodes développées soient rapides afin de pouvoir être embarquées. Les verrous scientifiques portent principalement sur la problématique 2, la "reconstruction 3D".
2. **Identification et suivi visuel** (responsable : Thierry Chateau, Maître de Conférence UBP). Cet axe vise à concevoir des méthodes de suivi et de reconnaissance des formes et vise des applications telles que la vidéo surveillance. Il porte principalement sur la problématique 1, la "mise en relation d'images".
3. **Vision en environnement déformable** (responsable : Adrien Bartoli, Chargé de Recherche CNRS). Nous visons dans cet axe à produire des méthodes de vision lorsque l'environnement observé se déforme. Les deux problématiques ci-dessus sont abordées : "mise en relation d'images" et "reconstruction 3D". L'accent est mis sur la modélisation des scènes déformables et l'utilisation de contraintes génériques.

Mes travaux sont répartis entre les axes 1 et 3 ci-dessus. Les travaux sur l'axe 1 sont une continuation directe de mes travaux de thèse, alors que ceux liés à l'axe 3 ont été initiés en partie lors de mon post-doc outre-Manche.

Il faut noter que l'axe 3 n'existait pas au LASMEA avant mon recrutement et que j'ai eu l'opportunité de pouvoir le démarrer. La structuration ci-dessus a été suggérée lors de l'évaluation quadriennale du LASMEA en 2006. Elle a été approuvée dans le rapport d'évaluation émis par ses tutelles.

L'équipe ComSee s'insère parfaitement dans la politique scientifique du site Clermontois, notamment au travers de la Fédération de Recherche TIMS. Cette dernière est structurée en 5 projets. ComSee s'insère au niveau des projets V2I (Véhicules et Infrastructures Intelligents), M2I (Machines et Mécanismes Innovants) et MLSVP (Modèles et Logiciels pour la Santé, le Vivant et le Physique). Les problématiques de l'axe 3 de ComSee font partie de celles affichées au niveau du projet MLSVP de TIMS. Notre axe 3 contribue de manière importante au projet MLSVP (réunions scientifiques, collaborations, co-encadrement d'un doctorant), et marginalement aux projets M2I (co-encadrement de stagiaires) et V2I (collaborations internes au GRAVIR). Des précisions scientifiques détaillées sont apportées au chapitre 3.

1.4 Encadrements et collaborations

J'indique mes encadrements de doctorants, post-doctorants et stagiaires, ainsi que mes collaborations avec d'autres laboratoires de recherche et chercheurs.

1.4.1 Doctorants

Je passe en revue ci-dessous les doctorants que j'encadre. Ils sont tous inscrit à l'Université Blaise Pascal de Clermont-Ferrand, sauf Pauline Julian qui est inscrite à l'Université Paul Sabatier de Toulouse. Florent Brunet est lui en co-tutelle avec la TUM en Allemagne. J'indique pour chaque doctorant l'année de démarrage de la thèse, le pourcentage d'encadrement que j'assure et les éventuels co-encadrants. Des références vers la description des travaux dans la partie II de ce document sont données.

- ▷ **Hanna Martinsson** (2004) co-encadrement (40%) avec François Gaspard du CEA Saclay, directeur de thèse : Jean-Marc Lavest. Sujet : *Reconstruction 3D d'objets manufacturés pour le contrôle qualité*. Hanna a travaillé sur la reconstruction 3D avec des caméras affines, puis s'est focalisée sur l'utilisation de courbes lisses comme primitives, à travers des approches basées primitive et basées pixel. Elle a commencé la rédaction de sa thèse en janvier 2008, après deux congés maternités. Nos travaux sont décrits en §§8.1.3, 8.3.1 et 8.3.2.
- ▷ **Mathieu Perriollat** (2005) co-encadrement (95%), directeur de thèse : Jean-Marc Lavest. Sujet : *Modélisation et reconstruction 3D d'objets de type papier*. Mathieu travaille sur la modélisation du papier par des surfaces déformables développables, et la reconstruction de ce modèle à partir d'images. Il a eu l'opportunité de faire des séjours scientifiques au DIKU (avec Søren Olsen), à la TUM (avec Nassir Navab) et à l'ANU (avec Richard Hartley). Il collabore avec Lionel Reveret de l'INRIA Grenoble. Mathieu a commencé la rédaction de sa thèse en janvier 2008. Nos travaux sont décrits en §§6.2.1, 7.1.5 et 7.2.1.
- ▷ **Vincent Gay-Bellile** (2005) co-encadrement (70%) avec Patrick Sayd du CEA Saclay, directeur de thèse : Jean-Thierry Lapresté. Sujet : *Suivi et reconstruction 3D de surfaces déformables*. Vincent travaille sur le suivi et la reconstruction 3D de surfaces déformables variées, comme le tissu, le papier, et les visages. Nous avons obtenu le **prix du meilleur article au congrès CORESA'07** pour nos travaux sur le suivi de surface avec prise en compte des auto-occultations. Vincent a eu l'opportunité de faire des séjours scientifiques au Queen Mary (Londres, avec Lourdes Agapito), à l'University d'Edimbourg (Ecosse, avec Bob Fisher) et au VIPS (avec Umberto Castellani). Il a commencé la rédaction de sa thèse en février 2008. Nos travaux sont décrits en §§6.2.4, 6.2.5, 7.1.4 et 7.1.5.
- ▷ **Julien Michot** (2007) co-encadrement (30%) avec François Gaspard du CEA Saclay, directeur de thèse : Jean-Marc Lavest. Sujet : *Auto-étalonnage de caméras embarquées*. Le but de la thèse de Julien est la conception d'un système d'auto-étalonnage de caméras embarquées par la technique décrite en §8.1.2.
- ▷ **Florent Brunet** (2007) co-encadrement (40%), directeurs de thèse en co-tutelle : Nassir Navab (TUM) et Rémy Malgouyres (LAIC, Université d'Auvergne (UdA)). Sujet : *Reconstruction 3D dense de surfaces déformables*. Nous désirons étudier des méthodes de reconstruction 3D de surfaces déformables à partir d'images d'une seule caméra. La reconstruction devra être dense. Elle sera utilisée dans le cadre médical interventionnel par endoscopie comme une aide à la perception pour le praticien. Nos premiers travaux sont décrits en §9.2.2.
- ▷ **Dawei Liu** (2007) co-encadrement (95%), directeur de thèse : Michel Dhome. Sujet : *Reconstruction 3D de surfaces déformables avec contraintes issues de la mécanique des milieux continus*. La plupart des algorithmes de suivi et de reconstruction de surfaces déformables monoculaires n'utilisent pas de contraintes sur la physique du matériaux considéré. Nous désirons exploiter de telles contraintes, issues de la mécanique des milieux continus, afin de rendre le problème mieux posé, et de réaliser la caractérisation du matériaux utilisé. Nous collaborons avec Michel Grédiac du LaMI (Laboratoire de Mécanique et Ingénieries, Clermont-Ferrand).

- ▷ **Samir Khoualed** (2007) directeur de thèse par dérogation. Sujet : *Ajustement de modèles génératifs à des vidéos*. L'objectif de cette thèse est l'ajustement de modèles génératifs à des vidéos en combinant les approches "bottom-up" et "top-down". Samir est actuellement en visite pour 6 mois au laboratoire VIPS (avec Umberto Castellani).
- ▷ **Pauline Julian** (2007) co-encadrement (10%) avec François Lauze (DIKU), directeur de thèse : Vincent Charvillat (IRIT). Sujet : *Méthodes variationnelles pour la segmentation et l'inpainting*. Pauline travaille dans le cadre d'un contrat CIFRE avec la société FittingBox. Le but est d'étudier comment les méthodes de segmentation et d'inpainting peuvent être utilisées pour l'augmentation d'images de visage.

1.4.2 Visiteurs

J'ai reçu la visite d'un certain nombre de chercheurs (Pierre Gurdjos, Umberto Castellani, Søren Olsen, Cristian Grava, ...) mais surtout d'étudiants inscrits en thèse. La plupart de ces visites ont données lieu à des publications.

- ▷ **Pierluigi Taddei** (6 mois, 2007-2008) étudiant à l'Université de Milan. Sujet : *Reconstruction 3D monoculaire de papier : une approche variationelle*.
- ▷ **Toby Collins** (2 mois, 2007) étudiant à l'Université d'Edimbourg. Sujet : *Recalage d'une surface déformable à partir de données 2,5D*. Notre algorithme est décrit en §7.2.3.
- ▷ **Julien Peyras** (2 mois, 2006 et 2007) étudiant à l'Université de Milan. Sujet : *Ajustement d'un modèle d'apparence actif de visage hiérarchique à une image*. Notre algorithme est décrit en §9.1.1. D'autres contributions communes sont décrites en §§9.1.2 et 7.1.4.
- ▷ **Benoît Bocquillon** (1 mois, 2006) étudiant à l'Université Paul Sabatier, Toulouse. Sujet : *Auto-étalonnage d'une caméra à distance focale constante*. Notre algorithme est décrit en §8.1.2.
- ▷ **Daniel Pizarro** (6 mois, 2006-2007) étudiant à l'Université de Madrid. Sujet : *Recalage d'images en dépit de variations d'éclairement et d'ombrage*. Notre algorithme est décrit en §6.1.3. Une autre contribution commune est décrite en §9.1.2.
- ▷ **Andreas Hofhauser** (1 mois, 2006) étudiant à l'Université Technique de Munich. Sujet : *Apprentissage artificiel d'un modèle de papier*.
- ▷ **Jean-Philippe Tardif** (2 mois, 2006) étudiant à l'Université de Montréal. Sujet : *Factorisation d'une matrice avec données manquantes et erronées*. Notre algorithme est décrit en §8.1.1.
- ▷ **Sylvie Chambon** (1 mois, 2006) étudiante à l'Université Paul Sabatier, Toulouse. Sujet : *Reconstruction 3D d'une plaque mince à partir de deux images*. Notre algorithme est décrit en §6.2.1.

1.4.3 Post-doctorants et stagiaires

1.4.3.1 Post-doctorants

- ▷ **Michela Farenzena** (1 an, 2007-2008) co-encadrement avec Youcef Mezouar (LASMEA). Sujet : *Navigation d'un drone par vision artificielle*. Michela travaille sur la conception d'un système de navigation d'un drone par vision artificielle robuste. Ceci inclut la détection et la gestion des singularités et une régularisation adaptative de la trajectoire.

1.4.3.2 Stagiaires en Master 2 Recherche

Les stagiaires de Master 2 Recherche que j'ai encadrés sont issus des Master VIRO (Vision et Robotique) ou MSIR (Modèles, Systèmes, Imagerie, Robotique) de l'Université Blaise Pascal, à l'exception de Nikolas Tiilikainen qui vient du DIKU.

- ▷ **Emilien Gaignette** (6 mois, 2008) co-encadrement (50%) avec Søren Olsen (DIKU). Sujet : *Suivi-par-détection d'une surface déformable lisse*. Emilien travaille sur la combinaison d'une méthode de suivi-par-détection basée sur des points et une fonction de coût basée pixel exploitant l'hypothèse d'une surface continue et lisse. Il part au DIKU 3 mois dans le cadre d'un échange Erasmus.
- ▷ **Nikolas Tiilikainen** (1 an, 2007-2008) co-encadrement (75%) avec Søren Olsen (DIKU). Sujet : *Suivi de brises de mer dans des images satellites MSG*. Nikolas est en Master au DIKU. Il effectue son stage de Master au LASMEA sous ma responsabilité, sur le suivi de brises de mer, dans le cadre du projet ANR STANDS-MSG. La méthode de suivi variationnelle que nous développons est basée sur la continuité temporelle et spatiale du front de brise de mer.
- ▷ **Pierre Petitprez** (6 mois, 2007) co-encadrement (50%) avec Vincent Lepetit (EPFL). Sujet : *Suivi et reconstruction 3D monoculaire de papier à partir de points clefs*. Pierre a travaillé sur une combinaison des méthodes de reconstruction 3D pour le papier proposées au LASMEA et pour les surfaces lisses proposées au CVLAB de l'EPFL, où il a passé 3 mois.
- ▷ **Manuel Grand-Brochier** (6 mois, 2007). Sujet : *Reconstruction 3D d'une surface déformable vue sur un arrière-plan statique*. L'idée de ce stage était d'exploiter la connaissance d'un arrière-plan statique pour calculer la pose de la caméra, permettant ensuite une "triangulation non-rigide" du premier plan déformable.
- ▷ **Muneeb Abid** (6 mois, 2007) co-encadrement (20%) avec François Berry (LASMEA). Sujet : *Implantation d'un suivi de plan sur une plate-forme hétérogène*. Le but de ce stage était de porter mon algorithme de suivi décrit en §6.1.2 pour le cas d'un plan, sur une plate-forme matérielle utilisant un FPGA et un DSP. La méthode est décrite dans [I44].
- ▷ **Ludovic Najac** (6 mois, 2005). Sujet : *Estimation de transformations plaque minces par points clefs*. L'idée de ce stage était la séparation des composantes rigides et déformables lors de l'estimation de transformations plaque minces entre points clefs.

1.4.3.3 Autres stagiaires

J'ai encadré ou co-encadré 9 stages de niveau inférieur au Master 2 depuis mon recrutement au CNRS en 2004, certains effectués par des binômes. La plupart étaient inscrit dans une formation de l'Université Blaise Pascal ou de l'Université d'Auvergne.

1.4.4 Collaborations

Je maintiens des échanges et collaborations actives avec plusieurs équipes de pointe du domaine de vision par ordinateur, au travers de visites croisées et de séjours d'étudiants. Voici les plus actifs de ces contacts : l'Image Group de la DIKU dirigé par Mads Nielsen (ma collaboration la plus active est avec Søren Olsen), la CAMPAR de la TUM dirigée par Nassir Navab, le VIPS de l'Université de Vérone dirigé par Vittorio Murino (je collabore principalement avec Umberto Castellani), le CVLAB de l'EPFL dirigé par Pascal Fua (je collabore avec Vincent Lepetit). Citons encore l'IRIT (principalement Pierre Gurdjos), le Queen Mary à Londres (Lourdes Agapito), l'Université d'Edimbourg (Bob Fisher), l'ANU (Richard Hartley) et l'INRIA Rhône-Alpes.

1.5 Enseignement

J'ai été moniteur à l'UJF durant ma thèse. J'enseigne maintenant principalement dans le Master 2 VIRO de l'UBP et au DIKU dans le cadre de mon "Visiting Professorship". Voici la liste de mes cours :

"Image Registration – 2D, 3D, Rigid and Deformable Scenes"
Cours Master, Université de Vérone 20h 2008
"Image Registration – 2D, 3D, Rigid and Deformable Scenes"

Cours Master et doctoral, DIKU, Copenhague	20h	2007
"3D Computer Vision"		
Cours Master, DIKU, Copenhague	20h	2006
Géométrie pour la vision		
Cours Master 2 Recherche, UBP, Clermont-Ferrand	4×14h	depuis 2004
Traitement d'image		
Cours Master 2 Recherche, UBP, Clermont-Ferrand	8h	2007
Optimisation numérique avec MATLAB		
Cours et TDe 4ème année, ENSCCF, Clermont-Ferrand	2×32h	2005 et 2006

J'ai par ailleurs assuré des interventions dans d'autres cours ou modules :

"Structure-from-Motion – 3D Feature and Camera Reconstruction"		
Cours Master, CAMPAR-TUM, Munich	2×1h30	2006 et 2007
Vision par ordinateur : un tour d'horizon		
Cours Master 2 Pro, UBP	3×4h	depuis 2005
Vision 3D non calibrée		
Cours doctoral, module Image, Université Montpellier II	2×3h	2005 et 2007
"C++ Coursework Module"		
TDe 3ème année, Université d'Oxford	10h	2004

1.6 Animation de la communauté scientifique

J'ai co-organisé les manifestations suivantes :

- ▷ **NORDIA 2008**, "Workshop on Nonrigid Shape Analysis and Deformable Image Analysis", Anchorage, Alaska, associé au congrès CVPR
  Co-organisation : Vincent Lepetit (EPFL), Alexander Bronstein (Technion), Michael Bronstein (Technion), Adrien Bartoli, Ron Kimmel (Technion) and Nassir Navab (TUM)
  Site web : *tosca.cs.technion.ac.il/nordia08*
- ▷ **DEFORM 2006**, "Workshop on Image Registration in Deformable Environments", Edimbourg, associé au congrès BMVC
  Co-organisation : Adrien Bartoli, Nassir Navab (TUM) et Vincent Lepetit (EPFL)
  Site web : *comsee.univ-bpclermont.fr/events/DEFORM06*
- ▷ **ORASIS 2005**, congrès des jeunes chercheurs en vision par ordinateur, Fournol, France
  Co-organisation : Thierry Chateau (LASMEA) et Adrien Bartoli
  Site web : *comsee.univ-bpclermont.fr/events/ORASIS05*

Par ailleurs, j'ai co-organisé un tutoriel au congrès ISMAR 2007 à Nara, Japon, intitulé "Computer Vision for AR – Rigid and Deformable Tracking Using Markers or Scene Features". Les co-organisateurs étaient Selim Benhimane (TUM), Vincent Lepetit (EPFL) et moi-même. Le site web est *campar.in.tum.de/ISMAR07TT*. Je suis intervenu durant 45 minutes sur le thème du suivi et de la reconstruction 3D de surfaces déformables.

Je participe à l'organisation du congrès 3DPVT 2008 qui se tiendra au Georgia Institute of Technology, Atlanta, USA, en juin 2008, en tant que "Publications Chair".

1.7 Participation à des comités de programme et activité de relecture

J'ai effectué des relectures d'articles pour les revues suivantes :

PAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence
TASE	IEEE Transactions on Automation Science and Engineering
TRO	IEEE Transactions on Robotics
TIP	IEEE Transactions on Image Processing
CVIU	Computer Vision and Image Understanding
JMIV	Journal of Mathematical Imaging and Vision
IVC	Image and Vision Computing
	The Annals of Statistics
CGForum	Computer Graphics Forum (Eurographics)
PRL	Pattern Recognition Letters
	IEE Proceedings – Vision, Image and Signal Processing
	Journal of the Electronics and Telecommunications Research Institute
JEI	Journal of Electronic Imaging (SPIE and IS&T)
DKE	Data and Knowledge Engineering
IJCA	Int'l Journal of Computers and Applications
JIAS	Journal of Image Analysis and Stereology
TS	Traitement du Signal

Je participe et ai participé à 19 comités de programme, listés ci-dessous :

ECCV	European Conf. on Computer Vision	2008	Marseille
AMDO	Int'l Conf. on Articulated Motion and Deformable Objects	2008	Andratx
VIIP	Int'l Conf. on Visualization, Imaging and Image Processing	2008	Palma de Mallorca
TWPJJ	Tribute Workshop to Peter Johansen	2008	Copenhagen
VISAPP	Int'l Conf. on Computer Vision Theory and Applications	2008	Madeira
ICCV	IEEE Int'l Conf. on Computer Vision	2007	Rio de Janeiro
CVPR	IEEE Int'l Conf. on Computer Vision and Pattern Recognition	2008	Anchorage
		2007	Minneapolis
CORESA	Compression et Représentation des Signaux Audiovisuels	2007	Montpellier
3DIM	Int'l Conf. on 3D Digital Imaging and Modeling	2007	Montréal
SCIA	Scandinavian Conf. on Image Analysis	2007	Copenhagen
BMVC	British Machine Vision Conf.	2008	Leeds
		2007	Warwick
		2006	Edinburgh
WDV	Workshop on Dynamical Vision	2007	Rio de Janeiro
		2006	Graz
AMI-ARCS	Augmented Environments for Medical Imaging...	2008	New York
		2006	Copenhagen
ICIP	IEEE Int'l Conf. on Image Processing	2008	San Diego
		2007	San Antonio
		2006	Atlanta
ORASIS	Congrès des jeunes chercheurs en vision par ordinateur	2007	Obernai
		2005	Fournol

J'ai par ailleurs été relecteur pour les congrès suivants :

ICRA	IEEE Int'l Conf. on Robotics and Automation	2005, 2008
ISMAR	IEEE and ACM Int'l Symposium on Mixed and Augmented Reality	2006, 2007
MICCAI	Int'l Conf. on Medical Image Computing and Computer Assisted Intervention	2007, 2008
CVPR	IEEE Int'l Conf. on Computer Vision and Pattern Recognition	2001 à 2006
ECCV	European Conf. on Computer Vision	2002, 2006
ICCV	IEEE Int'l Conf. on Computer Vision	2003

1.8 Expertises et consultances

J'ai été expert pour les Appels à Projets suivants de l'ANR :

RIAM	réseau pour la Recherche et l'Innovation en Audiovisuel et Multimédia	2005 – 2008
TechLog	Technologies Logicielles (ex RNTL)	2007
CSOSG	Concepts Systèmes et Outils pour la Sécurité Globale	2006 – 2008

Je suis consultant auprès de la société FittingBox issue de l'IRIT depuis 2007, sur des problématiques générales de vision en environnement déformable.

1.9 Participation à des projets de recherche

J'ai participé ou participe aux projets de recherche suivants :

- ▷ **SURF-3D – 3D Reconstruction of Deformable Surfaces from Endoscopic Images** (Projet PHC Procope, 2008-2010). Partenaires : TUM et LASMEA. Le but de ce projet est la conception d'un système de vision monoculaire endoscopique permettant la reconstruction 3D de la surface d'un organe pour aider le praticien en terme de positionnement.
- ▷ **3D Reconstruction of Deformable Surfaces by Integrating Mechanical Models** (Projet PHC Alliance, 2008-2010). Partenaires : LASMEA et Queen Mary (Université de Londres). Ce projet vise à l'intégration de contraintes issues de la mécanique au modèle de déformation de faible rang pour aider la reconstruction 3D monoculaire d'un environnement déformable.
- ▷ **SUN – Surface Unraveling with Application to Flexible Document Scanning** (Projet financé par l'ambassade de France au Danemark). Partenaires : LASMEA et DIKU. Le but de ce projet est de réaliser la mise à plat d'une surface déformable à partir d'images par une approche basée sur l'apprentissage artificiel.
- ▷ **CPER région Auvergne** Partenaires : LAIC, LIMOS et LASMEA. La collaboration sur la place Clermontoise entre ces trois laboratoires dans le cadre de la Fédération de Recherche TIMS porte entre autres sur le thème de la reconstruction et la mise à plat de surfaces à partir de données provenant de capteurs hétérogènes. Cette collaboration est soutenue par la Région Auvergne au travers du Contrat de Projet Etat Région et d'Innov@Pôle.
- ▷ **HFIBMR – High Fidelity Image-Based Modeling and Rendering** (Projet ANR "blanc", 2007-2010). Partenaires : WILLOW (ENPC / ENS / INRIA, Paris), LASMEA et ARTIS (INRIA Rhône-Alpes). On désire dans ce projet faire de la vision par ordinateur un outil pour la capture de modèles 3D haute précision, permettant de concurrencer en qualité de rendu les systèmes de vision active (utilisant par exemple un laser).

- ▷ **VIRAGO – Vision Rapide** (Projet ANR "jeunes chercheurs", 2007-2011). Partenaires : les trois équipes du groupe GRAVIR du LASMEA. On cherche à étudier une chaîne la plus complète possible permettant d'utiliser les caméras de type "rolling shutter" comme capteurs de vitesse instantanée. Des détails techniques sont donnés en §8.2.3.
- ▷ **Contrat de recherche** (2007-2009). Partenaires : LASMEA et CEA (Fontenay-aux-Roses). Le but de ce contrat est de transférer et d'adapter des techniques de localisation 3D par vision sur un drone.
- ▷ **Echange de chercheurs CNRS** (2007). Partenaires : LASMEA et ANU. Ce financement m'a permis de faire un séjour dans le laboratoire de Richard Hartley durant l'été 2007.
- ▷ **PMoCap – Computer Vision Based Motion Capture for Paper** (Projet "jeunes chercheurs" du GDR ISIS, 2007-2009). Partenaires : LASMEA et EVASION (INRIA Rhône-Alpes). Ce projet vise à utiliser nos travaux sur la modélisation et la reconstruction 3D du papier dans des systèmes de synthèse d'image, thème sur lequel l'équipe EVASION est spécialiste.
- ▷ **A Machine Learning Approach to Deformable Object Modeling** (Projet du CCUFB, 2007). Partenaires : TUM et LASMEA. L'objectif de ce projet était d'utiliser des techniques d'apprentissage artificiel afin de trouver un modèle de surface déformable simple à partir d'exemples générés synthétiquement.
- ▷ **STANDS-MSG – Spatio-Temporal Analysis of Deformable Structures in MSG Images** (Projet ANR "jeunes chercheurs", 2006-2009). Partenaires : COSTEL (Rennes), GREYC (Caen), Perception (INRIA Rhône-Alpes), LMD (Paris), LASMEA, VISTA (IRISA, Rennes). Ce projet vise à développer des méthodes de vision par ordinateur pour le traitement des images de type MSG (Météosat Seconde Génération). Notre tâche est le suivi des fronts de brise de mer.
- ▷ **AirPhoto** (2004-2005). Partenaires : LASMEA et l'entreprise "The Unkelbach Valley Software Work" (Allemagne). Le but de ce projet était l'intégration de contraintes multi-vues au logiciel AirPhoto. Nos résultats ont été communiqués lors d'une présentation invitée donnée au congrès IAAC'04, voir ci-dessous.
- ▷ **VISIRE - Vision-based 3D Reconstruction** (Projet Européen IST, 1999-2003). J'ai participé à ce projet durant ma thèse. Mon rôle a été la conception d'algorithmes de reconstruction 3D par ajustement de faisceaux.

1.10 Communications invitées

J'ai donné une présentation invitée à l'atelier du projet LIMA3D "Topics in Automatic 3D Modeling and Processing Workshop" à Vérone, en mars 2006, sur le thème du recalage d'images en présence d'un changement d'illumination, et j'ai participé à l'élaboration de la présentation invitée "New Solutions to an Old Problem : Multiple Image Registration With Sparse Ground Control Data" au congrès IAAC'04 (AARG Int'l Aerial Archaeology Conference) à Munich, avec Irwin Scollar et Rog Palmer. Finalement, j'ai été invité à donner une présentation de mes travaux sur la modélisation et le suivi en environnement déformable à la journée "Modélisation 3D" du GDR ISIS en novembre 2006, et de mes travaux sur le recalage d'images lors de la journée "Suivi visuel robuste en temps-réel".

J'ai visité plusieurs laboratoires pour des séjours de durées variables et j'ai eu l'occasion d'y donner des séminaires invités :

- ▷ "Deformable Image Registration and Generic Surface Reconstruction"
  2007 (novembre) IRIT, Toulouse
- ▷ "Parametric Methods for Registering Images of a Deforming Surface"
  2007 (août) ANU, Canberra
- ▷ "Feature-Driven Direct Non-Rigid Image Registration"
  2007 (mai) DIKU, Copenhague
- ▷ "Modeling and Reconstructing Paper From Multiple Images"

2007 (février) LAIC, Clermont-Ferrand
▷ "Feature-Driven Direct Non-Rigid Image Registration"
2007 (février) TUM, Munich
▷ "Vision in Deformable Environments – Two Case Studies"
2006 (décembre) EPFL, Lausanne
▷ "Groupwise Geometric and Photometric Direct Image Registration"
2006 (juillet) IRIT, Toulouse
▷ "Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model"
2006 (février) DIKU, Copenhague
▷ "Non-Rigid Alignments for Tracking and Augmenting Deformable Surfaces"
2006 (février) LTH, Lund
▷ "Non-Rigid Alignments for Tracking and Augmenting Deformable Surfaces"
2006 (février) IMM, Copenhague
▷ "Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model"
2006 (février) Université de Malmoe
▷ "Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model"
2006 (janvier) TUM, Munich
▷ "Towards Non-Rigid Structure-from-Motion"
2005 (novembre) Université Humboldt, Berlin
▷ "Non-Rigid Alignments for Tracking and Augmenting Deformable Surfaces"
2005 (septembre) INRIA Rhône-Alpes, Grenoble
▷ "Non-Rigid Alignments for Tracking and Augmenting Deformable Surfaces"
2005 (juillet) TUM, Munich
▷ La géométrie projective en vision 3D non calibrée
2005 (avril) Laboratoire de Mathématiques, Clermont-Ferrand
▷ Estimation directe d'alignements non-rigides
2004 (novembre) LASMEA, Clermont-Ferrand
▷ Estimation directe d'alignements non-rigides
2004 (avril) INRIA Rhône-Alpes, Grenoble
▷ Augmentation de séquences d'images de scènes non-rigides
2004 (janvier) INRIA Rhône-Alpes, Grenoble
▷ Augmentation de séquences d'images de scènes non-rigides
2004 (janvier) LASMEA, Clermont-Ferrand
▷ "Radial Basis Functions"
2003 (décembre) Université d'Oxford
▷ Paramétrisation pour la reconstruction 3D : points, droites, plans et caméras
2002 (novembre) LASMEA, Clermont-Ferrand

1.11 Responsabilités administratives et collectives

J'ai assumé ou assume les responsabilités suivantes :

2008	Membre de trois commissions de sélection en section 27 de l'IUT – UdA
2007-	Membre suppléant de la commission des relations internationales de l'UFR ST – UBP
2007-	Membre suppléant de la commission de spécialiste section 61 – UBP
2007-	Correspondant du projet MLSVP de la Fédération de Recherche TIMS au LASMEA
2007-	Responsable d'un partenariat Socrates-Erasmus avec le DIKU
2006-	Membre du conseil de laboratoire du LASMEA
2005	Organisateur des journées annuelles GRAVIR

2005-	Mise en place et administration du site web de l'équipe ComSee
2004-2007	Organisateur des séminaires du groupe GRAVIR
2001-2003	Organisateur des séminaires de vision à l'INRIA Rhône-Alpes
2003	Organisateur des journées de bienvenu des moniteurs de l'UJF, Grenoble
2001	Organisateur des journées annuelles de l'équipe Perception (INRIA Rhône-Alpes)

J'ai par ailleurs obtenu des bourses de voyage IEEE pour participer aux congrès CVPR 2001 (Hawaï), ICCV 2003 (Nice) et CVPR 2004 (Washington). Je suis intervenu lors de la fête de la science en 2002 à Grenoble et en 2007 à Clermont-Ferrand.

1.12 Publications

J'ai signé ou co-signé 82 communications scientifiques écrites, dont 33 issues de mes travaux de doctorat à l'INRIA et 49 issues de mes travaux de post-doctorat à l'Université d'Oxford et en tant que Chargé de Recherche CNRS au LASMEA. Les travaux rapportés dans ce document sont issus de ces 49 dernières publications. Celles-ci sont indiquées ci-dessous par une clef en gras, on trouve 4 revues (un PAMI, deux JMIV et un IVC), un article invité dans un atelier, 33 congrès ou ateliers internationaux et 10 congrès ou ateliers nationaux. On peut noter que je suis co-auteur sur 5 articles soumis à des revues par mes doctorants ou moi-même (deux PAMI, un CVIU et un IVC, et un IEEE Trans. on Neural Networks). J'ai par ailleurs co-edité deux actes, pour les journées ORASIS 2005 et l'atelier DEFORM 2006. Je donne ci-dessous une liste quasi-exhaustive de mes publications par catégories et ordre chronologique. La plupart peuvent être consultées sur ma page personnelle. J'indique pour chacune la section de ce document où elle est incluse.

1.12.1 Revues internationales (11)

Note: les publications [J05,J06,J07] sont issues de mes travaux de thèse mais ont été réalisées après la fin de cette dernière.

J12 **Groupwise Geometric and Photometric Direct Image Registration** §6.1.2
A. Bartoli
IEEE *Transactions on Pattern Analysis and Machine Intelligence*, accepted December 2007

J11 **Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation** §6.2.2
A. Bartoli
Journal of Mathematical Imaging and Vision, special issue: tribute to Peter Johansen, accepted December 2007

J10 **Implicit Non-Rigid Structure-from-Motion with Priors** §7.1.3
S. Olsen and A. Bartoli
Journal of Mathematical Imaging and Vision, special issue: tribute to Peter Johansen, accepted December 2007

J09 **Triangulation for Points on Lines** §8.2.2
A. Bartoli and J.-T. Lapresté
Image and Vision Computing, Vol. 26, No. 2, p. 315-324, February 2008

J08 **A Random Samping Strategy For Piecewise Planar Scene Segmentation**
A. Bartoli
Computer Vision and Image Understanding, Vol. 105, No. 1, p. 42-59, January 2007

J07 **Affine Approximation for Direct Batch Recovery of Euclidean Motion from Sparse Data**
N. Guilbert, A. Bartoli and A. Heyden
International Journal of Computer Vision, Vol. 69, No. 3, p. 317-333, September 2006

J06 **Structure-From-Motion Using Lines: Representation, Triangulation and Bundle Adjustment**
A. Bartoli and P. Sturm
Computer Vision and Image Understanding, Vol. 100, No. 3, p. 416-441, December 2005

J05 **The Geometry of Dynamic Scenes - On Coplanar and Convergent Linear Motions Embedded in 3D Static Scenes**
A. Bartoli
Computer Vision and Image Understanding, Vol. 98, No. 2, p. 223-238, May 2005

J04 **Motion Panoramas**
A. Bartoli, N. Dalal and R. Horaud
Computer Animation and Virtual Worlds, Vol. 15, No. 5, p. 501-517, Novembre 2004

J03 **The 3D Line Motion Matrix and Alignement of Line Reconstructions**
A. Bartoli and P. Sturm
International Journal of Computer Vision, Vol. 57, No. 3, p. 159-178, May/June 2004

J02 **Non-Linear Estimation of the Fundamental Matrix with Minimal Parameters**
A. Bartoli and P. Sturm
IEEE *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 4, p. 426-432, April 2004

J01 **Constrained Structure and Motion From Multiple Uncalibrated Views of a Piecewise Planar Scene**
A. Bartoli and P. Sturm
International Journal of Computer Vision, Vol. 52, No. 1, p. 45-64, April 2003

1.12.2 Articles invités (1)

V01 **Direct Image Registration With Gain and Bias** §6.1.1
A. Bartoli
Topics in Automatic 3D Modeling and Processing Workshop, Verona, Italy, March 2006

1.12.3 Congrès et ateliers internationaux (50)

I50 **Coarse-to-Fine Low-Rank Structure-from-Motion** §7.1.4
A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen and P. Sayd
CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

I49 **Light-Invariant Fitting of Active Appearance Models** §9.1.2
D. Pizarro, J. Peyras and A. Bartoli
CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

I48 **Automatic Quasi-Isometric Surface Recovery and Registration from 4D Range Data** §7.2.3
T. Collins, A. Bartoli and R. Fisher
BMVA *Symposium on 3D Video - Analysis, Display and Applications*, London, UK, February 2008

I47 **Deformable Surface Augmentation in Spite of Self-Occlusions**
V. Gay-Bellile, A. Bartoli and P. Sayd
ISMAR'07 - IEEE / ACM *Int'l Symposium on Mixed and Augmented Reality*, Nara, Japan, November 2007

I46 **Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning** §6.2.5
V. Gay-Bellile, A. Bartoli and P. Sayd
ICCV'07 - IEEE *Int'l Conf. on Computer Vision, Rio de Janeiro, Brazil*, October 2007

I45 **Adaptive Evolution of 3D Curves for Quality Control**
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
WISP'07 - IEEE *Int'l Symposium on Intelligent Signal Processing*, Alcalá, Spain, October 2007

I44 **Implementation of an Image Registration Algorithm on an Heterogeneous Platform**
M. Abid, S. Prasad Sah, F. Berry, F. Dias and A. Bartoli
ICDSC'07 - ACM / IEEE *Int'l Conf. on Distributed Smart Cameras*, PhD forum, Vienna, Austria, September 2007

I43 **Direct Image Registration with Adaptive Multi-Resolution**
C. Grava, A. Bartoli, V. Gay-Bellile, V. Buzuloiu and J.-M. Lavest
VVG'07 - *Workshop Vision, Video and Graphics* at BMVC'07, Warwick, UK, September 2007

I42 **Using Priors for Improving Generalization in Non-Rigid Structure-from-Motion**
S. Olsen and A. Bartoli
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007

I41 **Segmented AAMs Improve Person-Independent Face Fitting** §9.1.1
J. Peyras, A. Bartoli, H. Mercier and P. Dalle
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007

I40 **Feature-Driven Direct Non-Rigid Image Registration** §6.2.4
V. Gay-Bellile, A. Bartoli and P. Sayd
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007

I39 **An Adaptive Multi-Resolution Algorithm for Motion Estimation in Medical Image Sequences**
C. Grava, A. Bartoli, V. Gay-Bellile, V. Buzuloiu and J.-M. Lavest
ECCTD'07 - IEEE *European Conf. on Circuit Theory and Design*, Sevilla, Spain, August 2007

I38 **Joint Reconstruction and Registration of a Deformable Planar Surface Observed by a 3D Sensor** §7.2.4
U. Castellani, V. Gay-Bellile and A. Bartoli
3DIM'07 - *Int'l Conf. on 3D Digital Imaging and Modeling*, Montréal, Québec, Canada, August 2007

I37 **Energy-Based Reconstruction of 3D Curves for Quality Control** §8.3.2
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
EMMCVPR'07 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EZhou, Hubei, China, August 2007

I36 **A Quasi-Minimal Model for Paper-Like Surfaces** §7.2.1
M. Perriollat and A. Bartoli
BenCOS'07 - ISPRS *Int'l Workshop "Towards Benmarking Automated Calibration, Orientation, and Surface Reconstruction from Images"* at CVPR'07, Minneapolis, USA, June 2007

I35 **Kinematics From Lines in a Single Rolling Shutter Image** §8.2.3
O. Ait-Aider, A. Bartoli and N. Andreff
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

I34 **Generalized Thin-Plate Spline Warps** §6.2.1
A. Bartoli, M. Perriollat and S. Chambon
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

I33 **Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion** §8.1.1
J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert and S. Roy
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

I32 **On Constant Focal Length Self-Calibration From Multiple Views** §8.1.2
B. Bocquillon, A. Bartoli, P. Gurdjos and A. Crouzil
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

I31 **Shadow Resistant Direct Image Registration** §6.1.3
D. Pizarro and A. Bartoli
SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007

I30 **Reconstruction of 3D Curves for Quality Control** §8.3.1
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007

I29 **Image Registration by Combining Thin-Plate Splines with a 3D Morphable Model** §7.1.5
V. Gay-Bellile, M. Perriollat, A. Bartoli and P. Sayd
ICIP'06 - *Int'l Conf. on Image Processing*, Atlanta, GA, USA, October 2006

I28 **Groupwise Geometric and Photometric Direct Image Registration**
A. Bartoli
BMVC'06 - *British Machine Vision Conf.*, Edinburgh, UK, p. 157-166, Vol. I, September 2006

I27 **A Single Directrix Quasi-Minimal Model for Paper-Like Surfaces**
M. Perriollat and A. Bartoli
DEFORM'06 - *Workshop on Image Registration in Deformable Environments* at BMVC'06, Edinburgh, UK, p. 11-20, September 2006

I26 **Triangulation for Points on Lines**
A. Bartoli and J.-T. Lapresté
ECCV'06 - *European Conf. on Computer Vision*, Graz, Austria, p. 189-200, vol. III, May 2006

I25 **Towards 3D Motion Estimation from Deformable Surfaces** §7.2.2
A. Bartoli
ICRA'06 - IEEE *Int'l Conf. on Robotics and Automation*, Orlando, Florida, USA, May 2006

I24 **Feature-Based Estimation of Radial Basis Mappings for Non-Rigid Registration**
V. Charvillat and A. Bartoli
VMV'05 - *Int'l Fall Workshop on Vision, Modeling and Visualization*, Erlangen, Germany, p. 195-199, November 2005

I23 **Handling Missing Data in the Computation of 3D Affine Transformations** §8.1.3
H. Martinsson, A. Bartoli, F. Gaspard and J.-M. Lavest
EMMCVPR'05 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, Florida, USA, p. 90-106, November 2005

I22 **A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery** §7.1.2
A. Bartoli and S. Olsen
WDV'05 - *Workshop on Dynamical Vision* at ICCV'05, Beijing, China, October 2005

I21 **Estimating the Pose of a 3D Sensor in a Non-Rigid Environment**
A. Bartoli
WDV'05 - *Workshop on Dynamical Vision* at ICCV'05, Beijing, China, October 2005

I20 **On Aligning Sets of Points Reconstructed From Uncalibrated Affine Cameras**
A. Bartoli, H. Martinsson, F. Gaspard and J.-M. Lavest
SCIA'05 - *Scandinavian Conference on Image Analysis*, Joensuu, Finland, p. 531-540, June 2005

I19 **Euclidean Reconstruction Independent on Camera Intrinsic Parameters**
E. Malis and A. Bartoli
IROS'04 - IEEE / RSJ *Int'l Conf. on Intelligent Robots Systems*, Sendai, Japan, p. 2313-2318, October 2004

I18 **Direct Estimation of Non-Rigid Registrations** §6.2.3
A. Bartoli and A. Zisserman
BMVC'04 - *British Machine Vision Conf.*, London, UK, p. 899-908, vol. II, September 2004

I17 **Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences** §7.1.1
A. Bartoli, E. von Tunzelmann and A. Zisserman
CVPR'04 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Washington, DC, USA, p. 699-706, vol. I, June 2004

I16 **A Framework for Pencil-of-Points Structure-From-Motion** §8.2.1
A. Bartoli, M. Coquerelle and P. Sturm
ECCV'04 - *European Conf. on Computer Vision*, Prague, Czech Republic, p. 28-40, vol. II, May 2004

I15 **Towards Gauge Invariant Bundle Adjustment: A Solution Based on Gauge Dependent Damping**
A. Bartoli
ICCV'03 - IEEE *Int'l Conf. on Computer Vision*, Nice, France, p. 760-765, vol. II, October 2003

I14 **Multiple-View Structure and Motion from Line Correspondences**
A. Bartoli and P. Sturm
ICCV'03 - IEEE *Int'l Conf. on Computer Vision*, Nice, France, p. 207-212, vol. I, October 2003

I13 **Batch Recovery of Multiple Views with Missing Data using Direct Sparse Solvers**
N. Guilbert and A. Bartoli
BMVC'03 - *British Machine Vision Conf.*, Norwich, UK, p. 63-72, vol. I, September 2003

I12 **VISIRE. Photorealistic 3D Reconstruction from Video Sequences**
T. Rodriguez, P. Sturm, M. Wilczkowiak, A. Bartoli, M. Personnaz, N. Guilbert, F. Kahl, M. Johansson, A. Heyden, J. M. Menendez, J. I. Ronda and F. Jaureguizar
ICIP'03 - IEEE *Int'l Conf. on Image Processing*, Barcelona, Spain, September 2003

I11 **Motion from 3D Line Correspondences: Linear and Non-Linear Solutions**
A. Bartoli, R. Hartley and F. Kahl
CVPR'03 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, p. 477-484, vol. I, June 2003

I10 **From Video Sequences to Motion Panoramas**
A. Bartoli, N. Dalal, B. Bose and R. Horaud
MOTION'02 - IEEE *Workshop on Motion and Video Computing*, Orlando, USA, p. 201-207, December 2002

I09 **The Geometry of Dynamic Scenes - On Coplanar and Convergent Linear Motions Embedded in 3D Static Scenes**
A. Bartoli
BMVC'02 - *British Machine Vision Conf.*, Cardiff, UK, p. 394-403, September 2002

I08 **A Unified Framework for Quasi-Linear Bundle Adjustment**
A. Bartoli
ICPR'02 - IAPR *Int'l Conf. on Pattern Recognition*, Québec, Canada, p. 560-563, August 2002

I07 **On the Non-Linear Optimization of Projective Motion Using Minimal Parameters**
A. Bartoli
ECCV'02 - *European Conf. on Computer Vision*, Copenhagen, Denmark, p. 340-354, May 2002

I06 **Minimal Metric Structure and Motion from Three Affine Images**
M.-A. Ameller, A. Bartoli and L. Quan
ACCV'02 - *Asian Conf. on Computer Vision*, Melbourne, Australia, p. 356-361, January 2002

I05 **The 3D Line Motion Matrix and Alignment of Line Reconstructions**
A. Bartoli and P. Sturm
CVPR'01 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Hawaii, USA, p. 287-292, December 2001

I04 **Piecewise Planar Segmentation for Automatic Scene Modeling**
A. Bartoli
CVPR'01 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Hawaii, USA, p. 283-289, December 2001

I03 **Projective Structure and Motion from Two Views of a Piecewise Planar Scene**
A. Bartoli, P. Sturm and R. Horaud
ICCV'01 - IEEE *Int'l Conf. on Computer Vision*, Vancouver, Canada, p. 593-598, July 2001

I02 **Constrained Structure and Motion From N Views of a Piecewise Planar Scene**
A. Bartoli and P. Sturm
VAA'01 - *Int'l Symposium on Virtual and Augmented Architecture*, Dublin, Ireland, p. 195-206, June 2001

I01 **Structure and Motion from Two Uncalibrated Views Using Points on Planes**
A. Bartoli, P. Sturm and R. Horaud
3DIM'01 - *Int'l Conf. on 3D Digital Imaging and Modeling*, Québec, Canada, p. 83-90, June 2001

1.12.4 Congrès et ateliers nationaux (15)

N15 **Reconstruction de surface par validation croisée**
F. Brunet, A. Bartoli, R. Malgouyres et N. Navab
ROADEF'08 - *Journées de recherche opérationnelle et d'aide à la décision*, Clermont-Ferrand, France, février 2008

N14 **Recalage non-rigide direct avec prise en compte des auto-occultations au niveau image**
V. Gay-Bellile, A. Bartoli et P. Sayd
RFIA'08 - *congrès francophone de Reconnaissance des Formes et Intelligence Artificielle*, Amiens, France, janvier 2008

N13 **Gestion des occultations pour l'augmentation d'une surface déformable**
V. Gay-Bellile, A. Bartoli et P. Sayd
CORESA'07 - *Journées "COmpression et REprésentation des Signaux Audiovisuels"*, Montpellier, France, Novembre 2007
Awarded the best student paper prize

N12 **Estimation directe d'alignements non-rigides guidés par primitives**
V. Gay-Bellile, A. Bartoli et P. Sayd
ORASIS'07 - *Onzième congrès francophone des jeunes chercheurs en vision par ordinateur*, Obernai, France, Juin 2007

N11 **Modélisation et reconstruction de papier à partir de plusieurs images**
M. Perriollat et A. Bartoli
ORASIS'07 - *Onzième congrès francophone des jeunes chercheurs en vision par ordinateur*, Obernai, France, Juin 2007

N10 **Autocalibrage multi-vues d'une distance focale et mouvements critiques associés**
B. Bocquillon, A. Bartoli, P. Gurdjos et A. Crouzil
ORASIS'07 - *Onzième congrès francophone des jeunes chercheurs en vision par ordinateur*, Obernai, France, Juin 2007

N09 **Reconstruction de courbes 3D pour le contrôle de conformité**
H. Martinsson, F. Gaspard, A. Bartoli et J.-M. Lavest
ORASIS'07 - *Onzième congrès francophone des jeunes chercheurs en vision par ordinateur*, Obernai, France, Juin 2007

N08 **A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery**
A. Bartoli and S. Olsen
Danish Machine Vision Conference, Copenhagen, Denmark, p. 43-52, August 2006

N07 **A Single Directrix Quasi-Minimal Model for Paper-Like Surfaces**
M. Perriollat and A. Bartoli
Danish Machine Vision Conference, Copenhagen, Denmark, p. 43-52, August 2006

N06 **Alignement de reconstructions tridimensionnelles affines en présence de données images manquantes**
H. Martinsson, A. Bartoli, F. Gaspard et J.-M. Lavest
RFIA'06 - *Congrès francophone de Reconnaissance des Formes et Intelligence Artificielle*, Tours, France, janvier 2006

N05 **Des séquences vidéo aux panoramas de mouvement**
A. Bartoli, N. Dalal, B. Bose et R. Horaud
CORESA'03 - *Journées "COmpression et REprésentation des Signaux Audiovisuels"*, Lyon, France, p. 171-174, janvier 2003

N04 **Reconstruction métrique minimale à partir de trois caméras affines**
M.-A. Ameller, A. Bartoli et L. Quan
RFIA'02 - *Congrès francophone de Reconnaissance des Formes et Intelligence Artificielle*, Angers, France, p. 471-477, janvier 2002

N03 **La matrice de mouvement pour droites 3D, application à l'alignement de reconstructions de droites**
A. Bartoli et P. Sturm
RFIA'02 - *Congrès francophone de Reconnaissance des Formes et Intelligence Artificielle*, Angers, France, p. 29-37, janvier 2002

N02 **Segmentation en plans pour la modélisation automatique à partir d'images**
A. Bartoli
Journée "Coopération Analyse d'Image et Modélisation", Lyon, France, p. 37-40, juin 2001

N01 **Triangulation projective contrainte par multi-coplanarité**
A. Bartoli et P. Sturm
ORASIS'01 - *Congrès francophone des jeunes chercheurs en vision par ordinateur*, Cahors, France, p. 47-56, Juin 2001

1.12.5 Rapports de recherche (4)

R04 **Motion Panoramas**
A. Bartoli, N. Dalal and R. Horaud
INRIA Research Report 4771, Grenoble, France, March 2003

R03 **Euclidean Bundle Adjustment Independent of Camera Intrinsic Parameters**
E. Malis and A. Bartoli
INRIA Research Report 4377, Sophia, France, December 2001

R02 **Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters**
A. Bartoli and P. Sturm
INRIA Research Report 4236, Grenoble, France, August 2001

R01 **A Projective Framework for Structure and Motion Recovery From Two Views of a Piecewise Planar Scene**
A. Bartoli, P. Sturm and R. Horaud
INRIA Research Report 4070, Grenoble, France, October 2000

CHAPITRE

2

SYNTHÈSE DES TRAVAUX DE RECHERCHE

Mes contributions concernent le domaine de la vision par ordinateur. Quelques manuels et collections récents directement liés à mes travaux sont (Faugeras et al., 2001; Hartley and Zisserman, 2003), qui concernent la géométrie d'images multiples, et (Forsyth and Ponce, 2003; Paragios et al., 2005), qui abordent le domaine de manière globale.

Une des motivations pour la recherche en vision par ordinateur est l'omniprésence de l'image dans nos sociétés modernes. Ceci est dû au développement rapide d'ordinateurs puissants et de capteurs visuels à bas coût. En effet, les appareils photos numériques et les webcams fournissent maintenant des images et vidéos de bonne qualité. Ces capteurs sont petits, peuvent être facilement embarqués et ne sont pas invasifs. Ils entraînent une forte demande d'algorithmes et logiciels robustes de vision par ordinateur. La plupart des problèmes sur lesquels j'ai contribué ont des applications potentielles importantes. Par exemple, la reconstruction 3D en environnement rigide et déformable peut être utilisée en architecture pour la reconstruction d'immmeubles, dans l'industrie du film pour les effets spéciaux et en robotique pour la localisation. Le recalage d'images a des applications en réalité augmentée pour le changement de l'apparence ou l'augmentation d'une surface dans une vidéo et en imagerie médicale pour la fusion d'images multimodales, entre autres.

D'un autre côté, la recherche en vision par ordinateur est motivée par la curiosité intellectuelle, liée au problème de la perception et du raisonnement artificiels. La vision par ordinateur est un domaine de recherche de pointe, fortement lié à l'intelligence et à l'apprentissage artificiels. Ceci se retrouve dans certaines directions actuellement prises par la communauté scientifique, où l'apprentissage artificiel est de plus en plus utilisé dans des tâches telles que le suivi et la reconnaissance visuels. L'apprentissage artificiel est un domaine de recherche très actif. Quelques manuels récents sont (Bishop, 1995; Hastie et al., 2001).

Une image numérique est produite par un capteur. Elle résulte de l'interaction de la lumière avec la structure de la scène, qui peut être rigide ou déformable. L'étude des relations entre plusieurs images soulève deux problématiques principales, les deux problématiques transverses à l'équipe ComSee, énoncées en §1.3.3 : *mise en relation d'images* et *reconstruction 3D*. Les phénomènes mis en jeu lors de la formation d'une image sont hautement complexes. Ceci induit plusieurs questions :

▷ **Modélisation explicite et invariance.** Il serait extrêmement difficile de modéliser explicitement l'intégralité des phénomènes mis en jeu et d'en reconstruire les paramètres, ou de reconstruire la function plénoptique/[1] Il faut donc choisir si un phénomène doit être explicitement modélisé ou si la fonction d'observation doit être rendue invariante aux effets qu'il induit. Les phénomènes non modélisés entraînent la présence de données erronées (par rapport au modèle), pouvant être rejetées par des méthodes d'estimation robustes. Un exemple est celui des variations d'illumination en recalage d'images. Modéliser explicitement l'effet d'un changement global d'illumination est en général facile à l'aide de paramètres de gain et de bias sur la couleur des pixels, par exemple. Cependant, modéliser explicitement des changements d'illuminations complexes est très lourd, car cela nécessite de modéliser la structure de la scène, sa BRDF[2] et la position des sources de lumière. C'est un cas où l'invariance rend la modélisation beaucoup plus facile. Tout ceci est précisé en §2.2.1.
Nos choix suivent en général les règles suivantes : les changements d'illumination globaux sont modélisés explicitement, alors que les changements plus complexes sont gérés par invariance. La structure 3D de la scène est modélisée explicitement ou contenu dans un modèle déformable au niveau image. Dans le premier cas, une caméra est souvent aussi modélisée. Nous utilisons ce que nous appellons un guide statistique tel que les modèles "morphables" 3D (3DMM)[3] qui permettent notamment de contraindre la position des points de contrôle d'une déformation image. Les composants d'un modèle sont fortement liés à la nature de la fonction de coût utilisée, basée primitive ou basée pixel.

▷ **Généricité, connaissances a priori et sélection de la complexité.** Il est admis que plus de connaissances a priori (ou juste "a priori") sur un problème sont utilisées, meilleure est la solution obtenue. Il est difficile de trouver des a priori génériques, où générique signifie non spécifique à une classe d'objets. Un exemple typique d'a priori générique est celui du lissage spatial et temporel. Utiliser de tels a priori soulève le problème du réglage de leur influence sur l'estimation, modélisé par les *paramètres de lissage*. Une pondération trop forte sur ces a priori augmente le biais du résultat, alors qu'un poid trop faible en augmente la variance. C'est un problème d'apprentissage artificiel typique, qui se présente en estimation de modèles déformables image. Nous avons utilisé des outils comme la validation croisée pour estimer des paramètres de lissage, en particulier pour le problème d'estimation des déformations image en §2.2.2. L'idée est de minimiser une erreur de généralisation, décrivant la capacité d'un modèle à extrapoler à de nouvelles données. Le lecteur intéressé est renvoyé sur (Poggio et al., 2004) pour plus de détails sur la généralisation et la prédictivité en apprentissage artificiel.
Il y a différentes manières de mesurer la prédictivité d'un modèle. Nous avons utilisé la "Prediction Sum of Squares statistic" (Allen, 1974) et la validation croisée (Wahba and Wold, 1975). Les modèles de déformation génériques sont souvent empiriques et ne permettent pas de formuler une distribution paramétrique des résidus, ce qui exclut l'utilisation de nombreuses techniques de sélection de modèle.
Il y a au moins deux façon de changer la complexité d'un modèle. La première est d'ajouter ou d'enlever des "morceaux" au modèle, tels que des centres de déformation pour une transformation plaque mince. Ceci change le *nombre de paramètres libres*. Nous avons utilisé cette approche en §2.5.2 pour déterminer le nombre de centres de déformation d'une transformation plaque mince au travers de la "Prediction Sum of Squares statistic". L'autre possibilité pour changer la complexité d'un modèle est de changer le paramètre de lissage lors de l'estimation des paramètres du modèle. Ceci change le *nombre effectif de paramètres*, tel que définit dans (MacKay, 1992; Moody, 1992). C'est ce que nous faisons lorsque nous calculons le paramètre de lissage par validation croisée.

Organisation de ce chapitre. Nous abordons des sujets variés, de la reconstruction 3D rigide classique à l'estimation de modèles de déformation image. Les solutions que nous proposons sont basées sur des modèles et des outils que nous avons parfois améliorés, et qui sont souvent partagés par différents problèmes. Il y a donc

[1] La fonction plénoptique 7D caractérise le rayon lumineux observé à toute longueur d'onde, pour toute position et orientation de la caméra, et à chaque instant (Adelson and Bergen, 1991).

[2] "Bidirectional Reflectance Distribution Function" en Anglais – décrit comment une surface renvoie la lumière.

[3] "3D Morphable Models" en Anglais. Ces modèles contiennent une composante de forme et d'apparence.

au moins deux façons de voir mes contributions : soit par les aspects techniques (*i.e.* les méthodes et outils, par exemple la factorisation de matrice), soit par les buts (par exemple le recalage d'images). J'ai organisé le reste de ce chapitre de manière à ce que ces deux façons de voir soient présentes. J'ai suivi la première possibilité pour l'organisation interne en §2.1 et la deuxième possiblité en §§2.2, 2.3 et 2.4. Je donne une synthèse des modèles et méthodes d'estimation utilisées en §2.1. Mes contributions en recalage d'images sont données en §2.2. Elles sont souvent utilisées pour fournir des données d'entrée aux algorithmes de reconstruction 3D en environnement déformable ou rigide, respectivement résumés en §§2.3 et 2.4. Finalement, je présente d'autres travaux en §2.5.

L'aspect bibliographique est réduit à son strict minimum. Le lecteur pourra consulter l'exposé détaillé contenu dans les chapitres 4 à 9 de la partie II de ce document. Les mêmes notations sont utilisées, ce qui explique pourquoi les acronymes ne correspondent pas toujours aux expressions Françaises. Les notations sont introduites à mesure de l'exposé afin d'assurer la complétude du chapitre.

2.1 Modèles déformables et méthodes d'estimation

Note : La version Anglaise détaillée de cette section se trouve au chapitre 5 dans la partie II de ce document.

Le but de cette section est de donner un aperçu des outils que nous avons utilisés et auxquels nous avons contribués. Ces outils ont eu un impact transversal sur nos travaux. Comme pour la plupart des sciences pour l'ingénieur, il y a en général deux étapes principales lors de la résolution d'un problème de vision par ordinateur : une étape de modélisation et une étape d'estimation. L'étape de modélisation consiste à formuler un modèle mathématique décrivant le problème et les contraintes qui y sont associées, ainsi qu'une fonction de coût dont le minimum correspond à la solution recherchée. L'étape d'estimation consiste à calculer les paramètres du modèle à partir d'observations en minimisant la fonction de coût. Cette section présente deux classes d'outils de modélisation (§§2.1.1 et 2.1.2) et trois types de méthodes d'estimation (§§2.1.3, 2.1.4 et 2.1.5). Elle ne prétend pas être exhaustive, due notamment à la quantité d'algorithmes existants, et aux possibilités ouvertes par la littérature sur les techniques dédiées aux courbes et surfaces. Elle ne couvre pas les techniques de géométrie visuelle (par exemple les modèles de caméra et les tenseurs d'appariement).

Nous étudions deux notions clefs pour les modèles déformables : les *fonctions de déformation image* et les *guides de déformation statistiques*, abrégés respectivement par les termes *fonctions de déformation* et *guides*. Des exemples sont respectivement la fonction de déformation plaque mince (TPS)[4] de (Bookstein, 1989) et le modèle de forme statistique (SSM)[5] de (Cootes et al., 1991). Une fonction de déformation est habituellement guidée par des centres de déformation et met en correspondance les pixels de plusieurs images. Les guides incorporent des a priori et permettent de contrôler les fonctions de déformation. Ils peuvent tous deux être basés sur des entités 2D ou 3D. Voici la description de quelques utilisations de ces modèles :

- ▷ **Une surface déformable.** Une fonction de déformation 2D peut servir à recaler les images avec une approche basée pixel, comme montré en §2.2.2. Les phénomènes de discontinuité induits par exemple par les auto-occultations doivent être gérés. Si une reconstruction 3D est désirée, un guide 3D peut être utilisé, seul ou en conjonction avec une fonction de déformation 2D.
- ▷ **Un objet de classe connue.** Un guide 3D pré-appris permet de recaler les images et de trouver une reconstruction 3D, voir §2.2.2. L'apparence peut aussi être apprise, comme dans les modèles d'apparence actifs (AAM),[6] voir §2.5.1.
- ▷ **Un environnement déformable non structuré.** La littérature n'offre pas beaucoup de possibilités pour ce cas. Le modèle de faible rang (LRSM)[7] donne de bons résultats avec une approche basée primitive, voir §2.3.1.

[4] "Thin-Plate Spline" en Anglais.
[5] "Statistical Shape Model" en Anglais.
[6] "Active Appearance Model" en Anglais.
[7] "Low-Rank Shape Model" en Anglais.

Notations. Les scalaires sont habituellement en italique (par exemple j), les vecteurs en gras (par exemple $\mathbf{q}$) et les matrices en sans-sérif (par exemple P). La transposée, l'inverse et la pseudo-inverse sont notées comme dans respectivement $\mathbf{q}^\mathsf{T}$, A^{-1} et $\mathsf{A}^\dagger$. La norme deux d'un vecteur et la norme de Frobenius d'une matrice s'écrivent comme dans $\|\mathbf{u}\|_2$ et $\|\mathsf{A}\|_\mathcal{F}$.

2.1.1 Fonctions de déformation image 2D

Une fonction de déformation image 2D affecte à un point de l'image source le point correspondant dans l'image cible. De telles fonctions peuvent être obtenues de différentes manières, et notamment de manière constructive ou variationnelle. Ces deux approches sont très liées. Une propriété importante est que la fonction de déformation soit continue et dérivable, ou "lisse". Il est naturel de définir de telles fonctions comme solutions de problèmes variationnels avec un terme de donnée et un terme de lissage.

2.1.1.1 Généralités

Fonction de déformation paramétriques et principes généraux d'estimation. Soient $(\mathbf{q} \in \mathbb{R}^2) \leftrightarrow (\mathbf{q}' \in \mathbb{R}^2)$ une paire de points correspondants entre l'image source et l'image cible. Une *fonction de déformation paramétrique* $\mathcal{W} : \mathbb{R}^2 \times \mathbb{R}^p \mapsto \mathbb{R}^2$ à p paramètres s'écrit :

$$\mathbf{q}' \;=\; \mathcal{W}(\mathbf{q};\mathbf{u}).$$

Le *vecteur de paramètres* $\mathbf{u}$ peut contenir des quantités variées telles des points de contrôle ou les paramètres d'une surface et d'une caméra.

Le principe d'estimation par lissage est basé sur la minimisation d'une fonction de coût $\mathcal{E}_c$ composée d'au moins deux termes : le terme de donnée $\mathcal{E}_d$ et le terme de lissage externe $\mathcal{E}_s$. Le problème d'estimation paramétrique s'écrit donc :

$$\min_{\mathbf{u}} \mathcal{E}_c(\mathbf{u};\mu) \qquad \text{avec} \qquad \mathcal{E}_c(\mathbf{u};\mu) \;\stackrel{\text{def}}{=}\; \mathcal{E}_d(\mathbf{u}) + \mu\mathcal{E}_s(\mathbf{u}).$$

On note l'introduction du *paramètre de lissage* $\mu \in \mathbb{R}^+$, contrôlant l'influence du terme de lissage. L'estimation de ce paramètre par la minimisation d'une *erreur de généralisation* est une de nos contributions examinée en §2.1.3. L'utilisation des dérivées partielles de la fonction de déformation est très commune pour le terme de lissage. A l'ordre 2, avec $\mathsf{H}(\mathbf{q};\mathbf{u})$ la matrice Hessienne de la fonction de déformation évaluée au point $\mathbf{q}$ et pour les paramètres $\mathbf{u}$, cela donne :

$$\mathcal{E}^2_{s,2}(\mathbf{u}) \;\stackrel{\text{def}}{=}\; \sum_{\mathbf{q}\in\mathcal{R}} \|\mathsf{H}(\mathbf{q};\mathbf{u})\|^2_\mathcal{F} \qquad \text{avec} \qquad \mathsf{H}(\mathbf{q};\mathbf{u}) \;\stackrel{\text{def}}{=}\; \frac{\partial^2\mathcal{W}}{\partial\mathbf{q}^2}(\mathbf{q};\mathbf{u}).$$

Les termes de donnée les plus utilisés sont les termes basés pixel (ou "directs") et les termes basés primitive. Basé pixel signifie que la valeur (niveau de gris ou couleur) des pixels est directement comparée, typiquement par :

$$\mathcal{E}^2_{dp}(\mathbf{u}) \;=\; \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}))\|^2_2, \tag{2.1}$$

avec $\mathcal{S}$ l'image source, $\mathcal{T}$ l'image cible et $\mathcal{P}$ l'ensemble de pixels d'intérêt dans l'image source. De nombreuses autres solutions sont possibles, permettant l'invariance à des changements locaux ou globaux du signal image (corrélation croisée normée centrée, information mutuelle, ...). Les termes peuvent être rendus robustes afin que les phénomènes non modélisés, typiquement les occultations de la surface d'intérêt, ne viennent pas corrompre l'estimation. Nous utilisons un terme de donnée robuste avec un M-estimateur en §2.2.2. Les termes de donnée basés primitive utilisent des correspondances de primitives géométriques et sont souvent exprimés en pixels. Soient $(\mathbf{q}_j \in \mathbb{R}^2) \leftrightarrow (\mathbf{q}'_j \in \mathbb{R}^2)$ avec $j = 1,\ldots,m$ un ensemble de correspondances de points. L'*erreur de transfert* est un terme de donnée basé primitive typique qui s'écrit :

$$\mathcal{E}^2_{df}(\mathbf{u}) \;\stackrel{\text{def}}{=}\; \sum_{j=1}^{m} d^2(\mathbf{q}'_j, \mathcal{W}(\mathbf{q}_j;\mathbf{u})),$$

avec $d(\cdot,\cdot)$ la distance Euclidienne. Il est reconnu que les termes de donnée basés primitive permettent l'estimation de déformations de grande amplitude. La précision qu'ils permettent d'atteindre peut cependant laisser à désirer dans le cas des modèles déformables pour lesquels la redondance d'information présente dans l'image ne peut être exploitée que de manière locale. Les termes de donnée basés pixel sont en général plus précis, mais peuvent difficilement être directement utilisés pour des grandes déformations. Combiner les deux approches est un moyen efficace pour allier robustesse et précision.

Modéliser les variations photométriques des images est important pour les termes de donnée basés pixel, mais aussi pour des tâches comme l'incrustation dans une vidéo.

Le flot optique. Ce terme recouvre plusieurs notions. Nous l'utilisons comme le déplacement de l'ensemble des pixels de l'image source nécessaire pour transformer cette dernière vers l'image cible. En d'autres termes, le flot optique est une discrétisation de la fonction de déformation sur la grille des pixels.

2.1.1.2 Quelques fonctions de déformation paramétriques

Nous montrons que les deux fonctions de déformation les plus populaires, les déformation de forme libre (FFD)[8] et à base radiale (RBF),[9] au travers des TPSs, peuvent être obtenues à partir de la spline cubique 1D. L'idée est de combiner deux fonctions $\mathbb{R}^2 \mapsto \mathbb{R}$ partageant certaines propriétés afin de former la fonction $\mathbb{R}^2 \mapsto \mathbb{R}^2$ recherchée.

Il existe de nombreuses autres fonctions de déformation dans la littérature, comme les fonctions affines par morceaux, ainsi que de nombreuses possibilités pour en créer d'autres à partir d'outils provenants des domaines de synthèse d'image, du design assisté par ordinateur ou encore de la modélisation statistique de données.

La forme affine relevée. Nous définissons les fonctions de déformation *affines relevées* comme celles pouvant s'écrire sous la forme d'une projection $\mathbb{R}^l \mapsto \mathbb{R}^2$, via une matrice L inconnue, de coordonnées relevées non-linéaires avec fonction de "levage" $\nu : \mathbb{R}^2 \mapsto \mathbb{R}^l$ connue (LA est utilisé pour "Lifted Affine" en Anglais) :

$$\mathcal{W}_{\mathrm{LA}}(\mathbf{q};\mathsf{L}) \stackrel{\text{def}}{=} \mathsf{L}^{\mathsf{T}}\nu(\mathbf{q}).$$

Ce modèle général inclut les FFDs, les RBFs et donc les TPSs. L'écriture sous cette forme se fait par le guidage par primitives que nous proposons pour les TPSs. Nous montrons en §2.2.2 que cette forme s'étend à une forme *perspective relevée*, permettant de modéliser les effets de projection perspective dans les modèles déformables. Nous supposons ci-dessous que la matrice L est de taille $(l \times 2)$ et contient les points de contrôle cibles ou les centres de déformation cibles.

La spline cubique comme base de construction de fonctions de déformation lisses. Une spline est une fonction lisse polynomiale par morceaux. Ce nom a été introduit dans (Schoenberg, 1946). Considérons η points (x_k, z_k) donnés. La spline cubique $\psi : \mathbb{R} \mapsto \mathbb{R}$ est solution du problème variationnel suivant :

$$\min_{\psi} \int_{\mathbb{R}} \left(\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}\right)^2 \mathrm{d}x \qquad \text{t.q.} \qquad \psi(x_k) = z_k, \;\; k = 1,\ldots,\eta. \tag{2.2}$$

La fonctionnelle impliquée représente l'*énergie de torsion* de la courbe. Cette dernière est contrainte à passer par les points donnés. La solution reste la spline cubique si un terme d'attache aux données est ajouté à la fonctionnelle en remplacement des contraintes d'interpolation. L'extension de cette spline vers $\mathbb{R} \mapsto \mathbb{R}^d$ avec $d \geq 1$ est effectuée est remplaçant les scalaires z_k par des *points de contrôle cibles* dans $\mathbb{R}^d$.

En supposant que la distance inter-noeuds est l'unité, nous pouvons écrire la spline cubique comme :

$$\psi(x) \;\; = \;\; \sum_{a=0}^{3} B_a(x - \lfloor x \rfloor) z_{\lfloor x \rfloor + a},$$

[8] "Free-Form Deformation" en Anglais.
[9] "Radial Basis Function" en Anglais.

avec les fonctions de mélange :

$$B_0(x) \stackrel{\text{def}}{=} \tfrac{1}{6}(-x^3+3x^2-3x+1) \qquad B_1(x) \stackrel{\text{def}}{=} \tfrac{1}{6}(3x^3-6x^2+4)$$
$$B_2(x) \stackrel{\text{def}}{=} \tfrac{1}{6}(-3x^3+3x^2+3x+1) \qquad B_3(x) \stackrel{\text{def}}{=} \tfrac{1}{6}x^3.$$

Un des avantages de ces fonctions est leur support compact : la valeur en un point n'est influencée que par 4 points de contrôle voisins.

Les déformations de forme libre. Les FFDs, proposées dans (Sederberg and Parry, 1986), sont basées sur le produit tensoriel entre deux fonctions lisses $\mathbb{R} \mapsto \mathbb{R}$, souvent choisies comme des splines cubiques. Les points de contrôle source sont donc disposés sur une grille régulière. Considérons deux ensembles de noeuds avec comme précédemment une distance inter-noeuds unité, associés respectivement aux axes x et y. Les sommets de la grille régulière ainsi définie sont les *points de contrôle sources* de coordonnées $(u \;\; v)^{\mathsf{T}} \in \mathbb{N}^2$. La dernière étape consiste à associer à chacun d'eux un point de contrôle cible $\mathbf{c}_{u,v}$ au lieu d'une valeur scalaire cible $z_{u,v}$; ceci correspond à l'utilisation conjointe de deux fonctions $\mathbb{R}^2 \mapsto \mathbb{R}$ afin d'obtenir la fonction $\mathbb{R}^2 \mapsto \mathbb{R}^2$ recherchée. Pour un point $\mathbf{q} \in \mathbb{R}^2$ de coordonnées $\mathbf{q}^{\mathsf{T}} = (x \;\; y)$, la fonction produit tensoriel s'écrit :

$$\mathcal{W}_{\text{FFD}}(\mathbf{q}; \mathsf{L}) \stackrel{\text{def}}{=} \sum_{a=0}^{3}\sum_{b=0}^{3} B_a(x - \lfloor x \rfloor) B_b(y - \lfloor y \rfloor) \mathbf{c}_{\lfloor x \rfloor + a, \lfloor y \rfloor + b}.$$

Les FFDs sont des déformations affines relevées car elles s'écrivent sous la forme $\mathcal{W}_{\text{FFD}}(\mathbf{q}; \mathsf{L}) = \mathsf{L}^{\mathsf{T}} \nu_{\text{FFD}}(\mathbf{q})$, où les coordonnées relevées sont données par les 16 coefficients non nuls du produit tensoriel, arrangés de manière appropriée.

Il a été défini de nombreuses variantes et extensions de ces FFDs, dont les FFDs incrémentales, les FFDs hiérarchiques et les FFDs de Dirichlet, ces dernières relâchant la contrainte que les points de contrôle sources forment une grille régulière.

Les déformations plaque mince et à base radiale. Les TPSs ont été obtenues par (Duchon, 1976) comme solutions de l'extension au 2D du problème variationnel (2.2) :

$$\min_{\zeta} \sum_{k=1}^{l} (z_k - \zeta(\mathbf{b}_k))^2 + \lambda \int_{\mathbb{R}^2} \left\| \frac{\partial^2 \zeta}{\partial \mathbf{q}^2}(\mathbf{q}) \right\|_{\mathcal{F}}^2 \mathrm{d}\mathbf{q}. \tag{2.3}$$

Notons que comme dans le cas 1D le terme de donnée n'est pas obligatoire. Une preuve d'unicité fut établie par (Wahba, 1990), et ces résultats utilisés pour construire des fonctions de déformation par (Bookstein, 1989). La TPS s'écrit :

$$\varphi(\mathbf{q}; \boldsymbol{\xi}_{\mathbf{z};\lambda}) \stackrel{\text{def}}{=} \mathbf{a}^{\mathsf{T}} \tilde{\mathbf{q}} + \sum_{k=1}^{l} \varrho(\|\mathbf{q} - \mathbf{b}_k\|_2^2) w_k,$$

avec $\tilde{\mathbf{q}}^{\mathsf{T}} = (\mathbf{q}^{\mathsf{T}} \;\; 1)$. La fonction de base TPS pour la distance au carré est donnée par $\varrho(d^2) \stackrel{\text{def}}{=} d^2 \log(d^2)$. Les coefficients rassemblés dans le vecteur $\boldsymbol{\xi}_{\mathbf{z},\lambda}^{\mathsf{T}} \stackrel{\text{def}}{=} (\mathbf{w}^{\mathsf{T}} \;\; \mathbf{a}^{\mathsf{T}})$ sont calculés en résolvant un système linéaire. Les fonctions de déformation TPS s'écrivent en combinant deux TPSs avec centres de déformation coïncidants :

$$\mathcal{W}_{\text{TPS}}(\mathbf{q}; \Xi_{\mathsf{L},\lambda}) \stackrel{\text{def}}{=} \mathsf{A}\tilde{\mathbf{q}} + \sum_{k=1}^{l} \varrho(\|\mathbf{q} - \mathbf{b}_k\|_2^2) \mathbf{w}_k,$$

où $\Xi_{\mathsf{L},\lambda}$ contient les vecteurs de coefficients $\boldsymbol{\xi}_{\mathbf{z};\lambda}$ sur les axes x et y.

Le guidage par primitives que nous proposons permet d'écrire la TPS sous la forme $\tau(\mathbf{q}; \mathbf{z}, \lambda) = \varphi(\mathbf{q}; \boldsymbol{\xi}_{\mathbf{z},\lambda})$ avec :

$$\tau(\mathbf{q}; \mathbf{z}, \lambda) \stackrel{\text{def}}{=} \boldsymbol{\ell}_{\mathbf{q}}^{\mathsf{T}} \mathsf{X}_{\lambda} \mathbf{z}, \tag{2.4}$$

et $\boldsymbol{\ell}_{\mathbf{q}}^{\mathsf{T}} \stackrel{\text{def}}{=} (\varrho(\|\mathbf{q}-\mathbf{b}_1\|_2^2) \ \cdots \ \varrho(\|\mathbf{q}-\mathbf{b}_l\|_2^2) \ \tilde{\mathbf{q}}^{\mathsf{T}})$. Sous cette forme, la TPS dépend directement des centres de déformation cibles dans $\mathbf{z}$ ou L. La forme affine relevée est obtenue directement comme :

$$\mathcal{W}_{\text{TPS}}(\mathbf{q};\mathsf{L}) \ = \ \mathsf{L}^{\mathsf{T}}\nu_{\text{TPS}}(\mathbf{q}) \qquad \text{avec} \qquad \nu_{\text{TPS}}(\mathbf{q}) \stackrel{\text{def}}{=} \mathsf{X}_{\lambda}^{\mathsf{T}}\boldsymbol{\ell}_{\mathbf{q}}. \tag{2.5}$$

Un des avantages des TPSs est que les centres de déformation sources $\mathbf{b}_k$ peuvent être placés arbitrairement dans l'image, et notamment aux points de donnée. Le désavantage est que le support de la fonction de base ϱ est global. La TPS est une RBF si l'on omet sa partie affine paramétrée par $\mathbf{a}$. Il existe de nombreuses RBFs solutions de problèmes variationnels de la forme (2.3) avec différents termes de lissage, obtenues dans le cadre de certains espaces de Hilbert, les RKHS.[10] Certaines de ces RBFs ont un support local.

2.1.2 Le modèle de faible rang et autres guides statistiques

Les fonctions de déformation décrites ci-dessus peuvent avoir un très grand nombre de paramètres. Le but des guides statistiques est de réduire ce nombre de paramètres en intégrant des a priori reflétant les dépendances entre les points. Il existe de nombreux guides basés sur la physique. Nous nous intéressons ici aux guides multilinéaires pour leur flexibilité et leur grande capacité de représentation. Considérons tout d'abord que la classe de l'objet observé, par exemple celle des visages, est connue. Le SSM de (Cootes et al., 1991) est obtenu par Analyse en Composantes Principales (ACP) sur un ensemble de points clefs annotés sur des images d'apprentissage. Ce modèle est souvent ajusté à une seule image pour des tâches de détection, localisation ou encore segmentation. Son extension en 3D proposé au travers des 3DMMs permet de retrouver une forme 3D à partir d'une seule image (Blanz and Vetter, 1999). La limitation principale de ces guides est qu'ils nécessitent de connaître ce que contient la scène observée. Il a été récemment proposé le LRSM, qui permet de découvrir la structure des données (Bregler et al., 2000; Irani, 1999).

Nous distinguons deux caractéristiques importantes pour les guides multilinéaires :

- ▷ **Dimension : 2D ou 3D.** Un guide 3D combine un modèle de caméra avec une forme 3D déformable. Son avantage est qu'il contient directement la pose de la caméra et la structure 3D. Il est en revanche "plus non-linéaire" qu'un guide 2D.
- ▷ **Apprentissage : pré-appris ou non-appris.** Un guide pré-appris est dédié à une classe d'objets spécifique, alors qu'un guide non-appris s'adapte aux données. Les guides pré-appris sont plus stables et mieux posés mais moins génériques.

Le tableau ci-dessous résume quelques modèles multilinéaires :

	2D	3D
Pré-appris	SSM (modèle de forme statistique) linéaire	3DMM (modèle "morphable" 3D) au moins bilinéaire
Non-appris	LRSM (modèle de faible rang) implicite bilinéaire	LRSM (modèle de faible rang) explicite au moins trilinéaire

Il apparaît clairement que le modèle de faible rang explicite est le plus général, et inclut les autres modèles.

Un des buts du LRSM explicite est la reconstruction 3D monoculaire en environnement déformable, un problème en général mal-posé. L'utilisation d'a priori tels que le lissage spatial ou temporel permet d'obtenir des résultats visuellement significatifs.

2.1.2.1 Les guides pré-appris

L'apprentissage consiste à estimer les *formes de base* d'un guide, souvent à partir de données pré-alignées. Cette étape est équivalente à l'ajustement d'un guide non-appris à des données. Elle est cependant effectuée avec des données qui facilitent le processus, comme des visages numérisés en 3D dans (Blanz and Vetter, 1999).

[10] "Reproducing Kernel Hilbert Spaces" en Anglais.

Cas 3D : les modèles "morphables" 3D. La position des points 3D dans les 3DMMs est une combinaison linéaire de l formes de base $\mathbf{B}_{k,j} \in \mathbb{R}^3$ apprises avec des coefficients de forme $\alpha_k \in \mathbb{R}$. Les points image $\mathbf{q}_j \in \mathbb{R}^2$ sont obtenus par un opérateur de projection $\Pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$:

$$\mathbf{q}_j \;=\; \Pi\left(\sum_{k=1}^{l} \alpha_k \mathbf{B}_{k,j}\right). \tag{2.6}$$

Le modèle est donc trilinéaire si une caméra affine est utilisée. Nous utilisons ce modèle pour guider une fonction de déformation TPS en §2.2.2.

Cas 2D : les modèles de forme statistiques. Les SSMs combinent des formes de base 2D $\mathbf{b}_{k,j} \in \mathbb{R}^2$ apprises et obtiennent les points image avec l'équation bilinéaire $\mathbf{q}_j \;=\; \sum_{k=1}^{l} \alpha_k \mathbf{b}_{k,j}$. Nous avons utilisé les AAMs, obtenus en ajoutant une composante d'apparence aux SSMs, en §2.5.1.

2.1.2.2 Les guides non-appris

Contrairement aux guides appris, les guides non-appris doivent découvrir des structures et régularités dans un seul jeu de données. Ils sont utiles dans les environnements dynamiques et peu structurés, contenant des objets non identifiés.

Cas 3D : le modèle de faible rang explicite. Le LRSM explicite s'écrit comme le 3DMM (2.6), mais nécessite l'introduction d'un indice de vue i car il ne peut être ajusté à un seule image. Un point image $\mathbf{q}_{i,j} \in \mathbb{R}^2$ est donné par :

$$\mathbf{q}_{i,j} \;=\; \Pi_i\left(\sum_{k=1}^{l} \alpha_{i,k} \mathbf{B}_{k,j}\right), \tag{2.7}$$

Cas 2D : le modèle de faible rang implicite. Le LRSM implicite est obtenu à partir du LRSM explicite en rassemblant les coefficients de forme $\alpha_{i,k} \in \mathbb{R}$ et la partie rotationnelle des projections affines dans des matrices composites M_i. De manière similaire, les formes de bases sont rassemblées dans les $\mathbf{S}_j$. Ceci crée des matrices de projection $(2 \times r)$ implicites J_i et des vecteurs de forme $(r \times 1)$ implicites $\mathbf{K}_j$, avec $r = 3l$ le rang du modèle :

$$\mathbf{q}_{i,j} \;=\; \mathsf{J}_i \mathbf{K}_j + \mathbf{t}_i \qquad \text{avec} \qquad \mathsf{J}_i \overset{\text{def}}{=} \mathsf{M}_i \mathsf{E}^{-1} \qquad \text{et} \qquad \mathbf{K}_j \overset{\text{def}}{=} \mathsf{E}\mathbf{S}_j. \tag{2.8}$$

Notons la présence d'une "matrice de mélange" $(r \times r)$ E. Estimer les paramètres de ce modèle est la première étape dans l'estimation stratifiée du modèle explicite, comme décrit ci-dessous.

2.1.2.3 Utilisation de connaissances a priori

Il a été montré par plusieurs auteurs que le guide LRSM était d'autant mieux posé que des a priori étaient utilisés, voir notamment (Torresani et al., 2007) et notre contribution en §2.3.1. En particulier, le modèle est très sensible au nombre de formes de base ou du rang choisis. L'utilisation d'a priori réduit cette sensibilité. Nous proposons des a priori de lissage générique en §2.3.1.

2.1.2.4 Reconstruction 3D par faible rang

Il existe plusieurs approches pour l'estimation des LRSM, les modèles de faible rang non-appris :

▷ **L'approche stratifiée.** Initialement proposée dans (Bregler et al., 2000), cette approche est basée sur trois étapes. Dans la première, le LRSM implicite est estimé par factorisation d'une matrice de mesure. Nous avons contribué à cette étape par nos algorithmes présentés en §2.1.4. Nous avons montré comment calculer ce modèle à partir de correspondances de points et de courbes en §2.3.1. La deuxième étape est le calcul de la matrice de mélange, permettant de retrouver le LRSM explicite. La troisième étape est le raffinement non-linéaire par ajustement de faisceaux.

- ▷ **L'approche ACP probabiliste.** C'est une approche récente proposée dans (Torresani et al., 2007). Les coefficients de forme sont marginalisés grâce à un a priori Gaussien, proche dans l'idée d'une ACP probabiliste.
- ▷ **L'approche "coarse-to-fine".** C'est l'approche que nous proposons en §2.3.1, inspirée par le concept "Deformotion" (Yezzi and Soatto, 2003). L'idée est d'ajouter des formes de base au LRSM explicite jusqu'à ce qu'une erreur de généralisation, initialement décroissante, augmente. Chaque ajout est résolu par factorisation rang 1 d'une matrice. L'algorithme obtenu est très stable, et gère la projection perpective.

2.1.2.5 Sélection du nombre de formes de base

Sélectionner le nombre de formes de base revient à fixer la flexibilité du modèle, liée à sa complexité. Le fait que ce modèle soit empirique ne permet pas d'utiliser les critères classiques de sélection de modèle (AIC, BIC, GRIC et MDL), souvent exprimés de manière compacte grâce à l'hypothèse d'une distribution paramétrique des résidus. Une méthode basée sur les valeurs propres de la matrice de données est proposée dans (Yan and Pollefeys, 2006). Nous proposons en §2.3.1 d'utiliser la validation croisée, pour laquelle des détails sont donnés dans la section suivante.

2.1.3 La "Prediction Sum of Squares statistic" et la validation croisée

Choisir entre plusieurs modèles ou fixer le niveau de flexibilité d'un modèle est un problème très commun en vision par ordinateur et dans d'autres domaines. Dans le cas des modèles déformables, cela se traduit souvent par le choix du paramètre de lissage de la fonction de coût. Ceci est en général effectué par des moyens *ad hoc* comme plusieurs essais manuels avec inspection visuelle du résultat. Ce problème peut être vue comme de l'apprentissage artificiel. Nous utilisons une technique nommée "Prediction Sum of Squares statistic" (PRESS) introduite dans (Allen, 1974), très proche de la validation croisée (LOOCV)[11] due à (Wahba and Wold, 1975). Le PRESS et le LOOCV sont des mesures de la prédictivité d'un modèle. Le LOOCV dépend des paramètres de lissage. le PRESS est donc utilisé pour comparer différents modèles, et le LOOCV pour ajuster des paramètres de lissage. Nous avons utilisé la validation croisée pour le calcul du rang en factorisation non-rigide, pour l'ajustement d'un modèle de surface 2,5D, et l'estimation de modèles de déformation image affines relevés.

Calculer ces statistiques est souvent vue comme coûteux car l'idée clef est de tester le modèle sur chaque donnée, non incluse dans le jeu d'apprentissage. Il existe des approximations permettant de réduire le nombre de tests, comme la validation croisée avec v partitions. Il existe par ailleurs des formules non itératives, ne nécessitant pas d'estimer le modèle autant de fois qu'il y a de données. Ces formules existent pour le cas des moindres carrés linéaires. Elles sont exactes pour le PRESS, mais nous montrons en §2.2.2 qu'elles sont approximatives pour le LOOCV.

2.1.3.1 La "Prediction Sum of Squares statistic"

Soient $\mathbf{u}$ un vecteur de paramètres, f un modèle et $a_j \leftrightarrow b_j$ des données, avec $j = 1, \ldots, m$. Considérons un problème de moindres carrés non-linéaires avec la fonction de coût :

$$\mathcal{E}^2_{\text{NLS}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}) - b_j\right)^2 .$$

Soit $\mathbf{u}^\star_{\text{NLS},(j)}$ la solution de ce problème en ignorant la donnée j. Le PRESS est défini par :

$$\mathcal{K}^2_{\text{NLS}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}^\star_{\text{NLS},(j)}) - b_j\right)^2 .$$

[11] "Leave-One-Out Cross-Validation" en Anglais.

Considérons maintenant un problème de moindres carrés linéaires avec la fonction de coût :

$$\mathcal{E}^2_{\text{STD}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{a}_j^\mathsf{T}\mathbf{u} - b_j\right)^2 = \frac{1}{m}\|\mathsf{A}\mathbf{u} - \mathbf{b}\|_2^2.$$

La formule non itérative du PRESS est :

$$\mathcal{K}^2_{\text{STD}} = \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1} - \text{diag}\left(\hat{\mathsf{A}}\right)}\right)\left(\hat{\mathsf{A}} - \mathrm{I}\right)\mathbf{b}\right\|_2^2,$$

où $\hat{\mathsf{A}} = \mathsf{A}\mathsf{A}^\dagger$ est la *matrice chapeau* et diag produit une matrice diagonale à partir d'un vecteur et extrait un vecteur contenant la diagonale d'une matrice, comme en MATLAB.

Le score du PRESS est typiquement minimisé en essayant tous les modèles disponibles. Lorsque ceci n'est pas possible, comme dans le cas des fonctions de déformation image, nous démarrons avec un modèle très simple, avec peu de centres de déformation, et insérons des centres jusqu'à ce que le PRESS augmente. Plus de détails sont donnés en §2.5.2. Cette formule ne s'applique pas directement dans le cas des problèmes homogènes (sans membre droit).

2.1.3.2 La validation croisée

Considérons maintenant un problème régularisé avec la fonction de coût :

$$\mathcal{E}^2_{\text{RNLS}}(\mathbf{u};\mu) \stackrel{\text{def}}{=} \mathcal{E}^2_{\text{NLS}}(\mathbf{u}) + \mu^2\mathcal{E}^2_s(\mathbf{u}),$$

avec $\mathcal{E}_s$ le terme de lissage. Le LOOCV est défini comme le PRESS, mais dépend du paramètre de lissage μ :

$$\mathcal{G}^2_{\text{RNLS}}(\mu) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}^\star_{\text{RNLS},(j)}(\mu)) - b_j\right)^2.$$

Considérons maintenant le cas linéaire, avec un terme de lissage de la forme $\mathcal{E}^2_s(\mathbf{u}) \stackrel{\text{def}}{=} \|\mathsf{Z}\mathbf{u}\|_2^2$. La matrice chapeau est remplacée par la *matrice d'influence* $\mathsf{T}(\mu) \stackrel{\text{def}}{=} \mathsf{A}\left(\mathsf{A}^\mathsf{T}\mathsf{A} + m\mu^2\mathsf{Z}^\mathsf{T}\mathsf{Z}\right)^{-1}\mathsf{A}^\mathsf{T}$, et le LOOCV est approximé par :

$$\mathcal{G}^2_{\text{RSTD}}(\mu) \approx \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1} - \text{diag}\left(\mathsf{T}(\mu)\right)}\right)\left(\hat{\mathsf{A}} - \mathrm{I}\right)\mathbf{b}\right\|_2^2.$$

La validation croisée généralisée de (Wahba, 1990) est basée sur l'approximation supplémentaire $\text{diag}(\mathsf{T}(\mu)) \approx \text{tr}(\mathsf{T}(\mu))\mathrm{I}$, avec I la matrice identité.

Minimiser le LOOCV sur le paramètre de lissage μ n'est pas un problème facile. La fonction $\mathcal{G}_{\text{RSTD}}$ a souvent une forme convexe, mais rien ne le garantit formellement, et il existe des cas pathologiques avec plusieurs minima locaux. Il est courant d'échantillonner la fonction, tout en combinant avec des descentes de gradient. Nous utilisons une méthode simplex qui donne de bons résultats, comme montré en §2.2.2.

2.1.4 Factorisation d'une matrice avec données manquantes et erronées

De nombreux algorithmes en vision par ordinateur et dans d'autres domaines nécessitent de factoriser une matrice. Citons la réduction de dimension, l'ACP, le "collaborative filtering", la reconstruction 3D, la reconstruction basée illumination, la segmentation par le mouvement ou encore la séparation du style et du contenu. Lorsque la matrice de mesure M est complète et sans erreur, la solution est obtenue par décomposition en valeurs singulières (SVD)[12], voir par exemple (Srebro and Jaakkola, 2003). Les données manquantes et erronées sont souvent inévitables en pratique, et rendent le problème beaucoup plus difficile. La plupart des algorithmes sont itératifs. Les algorithmes que nous proposons utilisent uniquement des étapes d'optimisation convexe.

[12] "Singular Value Decomposition" en Anglais.

2.1.4.1 Formulation du problème

Nous notons W une matrice binaire indiquant les données manquantes dans la matrice de mesure M de taille $(n \times m)$. Le problème est de trouver deux facteurs A et B de taille respective $(n \times r)$ et $(r \times m)$, où r est le rang de factorisation souhaité :

$$\min_{\mathsf{A},\mathsf{B}} \|\rho(\mathsf{W} \odot (\mathsf{M} - \mathsf{AB}))\|_{\mathcal{F}}^2 ,$$

avec $\odot$ le produit d'Hadamard (terme à terme), et ρ un M-estimateur (par terme). Il existe d'autres manières d'intégrer l'aspect robuste. Nos algorithmes n'utilisent pas de M-estimateur mais RANSAC.

2.1.4.2 Les contraintes de fermeture et de base

Les contraintes de fermeture sont dues à (Triggs, 1997b). Nous proposons les contraintes de base dans §2.4.1. Ces contraintes permettent de trouver un des deux facteurs en éliminant l'autre. Sans perte de généralité, nous calculons le premier facteur, c'est-à-dire A, en premier. Le deuxième facteur est ensuite trivial à estimer par moindres carrés linéaires.

Les contraintes de fermeture. Considérons une version $\mathcal{M}$ non bruitée et sans erreur des données. Une sous matrice de taille $(\tilde{n} \times \tilde{m})$ complète $\tilde{\mathcal{M}} = \mathsf{\Pi}\mathcal{M}\mathsf{\Gamma}$ de rang au moins r est sélectionnée en supprimant des lignes et des colonnes par les matrices binaires $\mathsf{\Pi}$ et $\mathsf{\Gamma}$. Elle est factorisée par SVD en $\tilde{\mathcal{M}} \rightarrow \mathsf{U\Sigma V}^\mathsf{T}$. Les $\tilde{m} - r$ dernières colonnes de U notées $\tilde{\mathcal{N}}$ forment une base pour le noyau gauche de $\tilde{\mathcal{M}}$, *i.e.* $\tilde{\mathcal{N}}^\mathsf{T}\tilde{\mathcal{M}} = 0$. Dans le cas de données bruitées, $\tilde{\mathcal{N}}$ est la meilleure approximation au sens des moindres carrés. Soit $\mathcal{M} = \mathcal{AB}$ la factorisation recherchée, nous obtenons $\tilde{\mathcal{N}}^\mathsf{T}\mathsf{\Pi}\mathcal{AB}\mathsf{\Gamma} = 0$. Comme $\mathsf{\Pi}\mathcal{A}$ et $\mathcal{B}\mathsf{\Gamma}$ sont de rang r, tout élément dans l'espace engendré par les colonnes de $\tilde{\mathcal{N}}$ est dans le noyau gauche de $\mathsf{\Pi}\mathcal{A}$, ce qui donne la *contrainte de fermeture sur le premier facteur* :

$$\mathcal{N}^\mathsf{T}\mathcal{A} = 0 \qquad \text{avec} \qquad \mathcal{N}^\mathsf{T} \stackrel{\text{def}}{=} \tilde{\mathcal{N}}^\mathsf{T}\mathsf{\Pi}.$$

La matrice $\mathcal{N}$ est un tenseur d'appariement dans le cas où cet algorithme est employé pour la reconstruction 3D. Cette matrice est souvent très éparse.

Les contraintes de base. Reprenons la SVD de la sous matrice complète $\tilde{\mathcal{M}}$. Les contraintes de fermeture sont basées sur le noyau gauche de cette matrice, mais ignorent les r premières colonnes $\bar{\mathsf{U}}$ du facteur U, donnant une base orthonormale de $\tilde{\mathcal{M}}$. La contrainte de base sur le premier facteur est formulée à l'aide d'une matrice $(r \times r)$ d'alignement Z :

$$\mathsf{\Pi}\mathcal{A} = \bar{\mathsf{U}}\mathsf{Z}.$$

Ces contraintes sont duales aux contraintes de fermeture car elles génèrent une base des inconnues au lieu d'exprimer directement les contraintes. Elles correspondent à une reconstruction 3D partielle exprimée dans une base qui lui est propre.

Résolution. Les contraintes de fermeture ou de base provenant de différentes sous matrices complètes sont combinées. Cela donne un système de moindres carrés linéaires dont la solution est en général une très bonne approximation du facteur A recherché.

2.1.4.3 Recherche de sous matrices complètes

La recherche de sous matrices complètes est une étape clef dans nos algorithmes. Il faut en trouver suffisamment pour que le premier facteur soit bien contraint. Chaque ligne de la matrice de mesure doit être impliquée dans au moins r sous matrices. Notre algorithme est basé sur une distribution aussi homogène que possible des contraintes le long des lignes. Nous passons séquentiellement en revue les colonnes de la matrice de mesure, et sélectionnons aléatoirement un certain nombre de lignes (les lignes et colonnes d'une sous matrice ne sont pas nécessairement contiguës). Un compteur est associé à chaque ligne, et permet de trouver facilement les lignes impliquées dans peu de contraintes, ainsi que d'en assurer une bonne répartition.

2.1.4.4 Robustification

Nos algorithmes peuvent être rendus robustes à trois niveaux. Le premier est lors de la factorisation par SVD des sous matrices. Il est facile de sélectionner les colonnes cohérentes par RANSAC. Ceci est équivalent à calculer par exemple une matrice fondamentale en sélectionnant les bonnes correspondances de points. Le deuxième niveau de robustification est lorsque les contraintes émanant de plusieurs sous matrices sont combinées. Un schéma de moindres carrés itérativement repondérés peut par exemple être utilisé. Finalement, l'estimation du deuxième facteur est réalisée par RANSAC. Cela correspond à une étape de triangulation dans l'analogie avec la reconstruction 3D.

Nous avons appliqué ces algorithmes avec grand succès aux problèmes de la reconstruction 3D en §2.4.1 et de la factorisation non-rigide en §2.3.1.

2.1.5 Recalage d'images compositionnel

Le recalage d'images est souvent effectué au travers de la minimisation d'une fonction de coût de moindres carrés non-linéaires basée primitive ou basée pixel. Les méthodes itératives telles Gauss-Newton avec convergence théorique superlinéaire donnent de très bons résultats à partir d'une solution initiale décente. Citons deux exemples : l'algorithme de Lucas-Kanade (Lucas and Kanade, 1981) et l'ajustement de faisceaux (Triggs et al., 2000). Ces méthodes linéarisent localement les termes de la fonction de coût, ce qui conduit aux équations normales, dont la solution donne à chaque itération la mise à jour du vecteur de paramètres. La matrice de coefficients des équations normales varie avec les itérations. Elle doit donc être calculée et inversée à chaque itération.

Il a été montré dans (Baker and Matthews, 2004) que sous certaines hypothéses cette matrice est constante. En d'autres termes, seul le vecteur membre droit des équations normales doit être recalculé à chaque itération. Ceci est possible grâce aux lois de mise à jour compositionnelles.

Soit $\mathbf{u} \in \mathbb{R}^p$ le vecteur de paramètres à estimer. Il est classique d'utiliser une loi de mise à jour additionelle, s'écrivant $\mathbf{u} \leftarrow \mathbf{u} + \boldsymbol{\delta}$, avec $\boldsymbol{\delta}$ le vecteur de mise à jour. Les lois de mise à jour compositionnelles directe et inverse s'écrivent respectivement :

$$\mathcal{W}(\cdot;\mathbf{u}) \leftarrow \mathcal{W}(\mathcal{W}(\cdot;\boldsymbol{\delta});\mathbf{u}) \qquad \text{et} \qquad \mathcal{W}(\cdot;\mathbf{u}) \leftarrow \mathcal{W}(\mathcal{W}^{-1}(\cdot;\boldsymbol{\delta});\mathbf{u}).$$

Ces lois nécessitent respectivement une structure de semi-groupe et de groupe sur la fonction de déformation. Leur utilisation initiale en recalage est due à (Shum and Szeliski, 2000). Partons du terme de donnée basé pixel (2.1) à minimiser sur les paramètres $\mathbf{u}_g$ d'une fonction de déformation :

$$\min_{\mathbf{u}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))\|_2^2.$$

L'introduction de la loi de composition inverse, et le basculement de la transformation incrémentale $\mathcal{W}^{-1}(\cdot;\boldsymbol{\delta})$ de l'image cible à l'image source, combinées avec un développement de Gauss-Newton nous donne le problème de moindres carrés linéaires suivant :

$$\min_{\boldsymbol{\delta}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) + \mathsf{L}_g^\mathsf{T}(\mathbf{q})\boldsymbol{\delta}_g - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))\|_2^2 \qquad \text{avec} \qquad \mathsf{L}_g^\mathsf{T}(\mathbf{q}) \stackrel{\text{def}}{=} (\nabla\mathcal{S})(\mathbf{q})^\mathsf{T}(\nabla_{\mathbf{u}_g}\mathcal{W})(\mathbf{q};\mathbf{0}).$$

Les matrices Jacobiennes L_g sont constantes, ce qui permet de "pré-résoudre" le système.

L'algorithme a trois étapes principes : *(i)* l'image cible est transformée vers l'image source avec les paramètres courants, *(ii)* un recalage "local" est effectué et *(iii)* les paramètres sont mis à jour.

Nous proposons en §2.2.1 la *composition inverse duale*. Cela permet de calculer efficacement les paramètres géométriques et photométriques du recalage. Les autres algorithmes perdent l'efficacité de la composition inverse ou nécessitent des approximations qui rendent l'optimisation peu performante (Baker et al., 2003). Soit $\mathcal{V}$ une transformation photométrique avec paramètres $\mathbf{u}_p$ portant sur la couleur des pixels. Notre loi de composition photométrique inverse est :

$$\mathcal{V}(\cdot;\mathbf{u}_p) \leftarrow \mathcal{V}^{-1}(\mathcal{V}(\cdot;\mathbf{u}_p);\boldsymbol{\delta}_p),$$

ce qui donne une loi complète :

$$\mathcal{V}(\mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g));\mathbf{u}_p) \leftarrow \mathcal{V}^{-1}(\mathcal{V}(\mathcal{T}(\mathcal{W}(\mathcal{W}^{-1}(\mathbf{q};\boldsymbol{\delta}_g);\mathbf{u}_g));\mathbf{u}_p);\boldsymbol{\delta}_p). \tag{2.9}$$

Ces lois permettent d'estimer des transformations affines mixant les canaux couleurs simultanément avec par exemple une homographie. Nous décrivons une méthode permettant d'appliquer le schéma de composition inverse à des fonctions de déformation sans structure de groupe et une méthode de recalage local basée apprentissage artificiel en §2.2.2.

2.2 Recalage d'images

Note : La version Anglaise détaillée de cette section se trouve au chapitre 6 dans la partie II de ce document.

Nous étudions dans cette section le recalage d'images en 2D ou, de manière équivalente, le calcul d'une fonction de déformation entre deux images. La première partie présente trois articles sur les aspects photométriques, cruciaux pour les approches basées pixel. La deuxième partie concerne cinq articles focalisés sur l'aspect environnement déformable.

2.2.1 La photométrie en recalage basé pixel

V01 **Direct Image Registration With Gain and Bias**
A. Bartoli
Topics in Automatic 3D Modeling and Processing Workshop, Verona, Italy, March 2006

J12 **Groupwise Geometric and Photometric Direct Image Registration**
A. Bartoli
IEEE *Transactions on Pattern Analysis and Machine Intelligence*, accepted December 2007
Version antérieure : [I28]
Article connexe : [I44]

I31 **Shadow Resistant Direct Image Registration**
D. Pizarro and A. Bartoli
SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007

Les deux premiers articles sont inspirés par les travaux de (Baker et al., 2003), qui propose des algorithmes permettant d'étendre le cadre inverse compositionnel présenté en §2.1.5 aux transformations photométriques. Ces algorithmes sont très généraux, mais sont en pratique lents ou peu fiables.

Le premier article propose une méthode *ad hoc* permettant d'estimer un changement global d'illumination, modélisé par un gain et un biais, entre deux images en niveau de gris. Une loi de composition inverse est utilisée sur les paramètres géométriques, et un mécanisme de résolution emprunté à l'ajustement de faisceaux permet de ne pas avoir à inverser de matrice à chaque itération. Cette approche est très performante, mais ne s'étend pas aux images couleurs, ni à des modèles photométriques plus complets.

Le deuxième article propose une méthode plus générale, adaptée aux images couleurs et à des changements photométriques globaux, par exemple, la recombinaison des canaux de couleurs. L'algorithme est basé sur notre loi de composition inverse duale (2.9). L'image cible est utilisée comme génératrice de l'image source. L'approche est rapide et stable. La méthode ne s'étend cependant pas aux changements de photométrie non globaux, comme ceux introduits par les ombres portées.

Le troisième article présente une méthode basée sur les espaces invariants à l'illumination de (Finlayson et al., 2002). L'idée est de "projeter" l'image couleur dans un espace 1D invariant. Les hypothèses sont que la lumière est Planckienne et la surface observée Lambertienne. Ceci n'est jamais vraiment vérifié en pratique, mais les résultats obtenus sont cependant satisfaisants. Cette théorie a été utilisée pour ôter les ombres d'une image. Nous proposons de l'utiliser pour le recalage. Notre idée est de calculer le coût entre les projections des deux images à recaler dans l'espace 1D invariant. La projection dépend de certains paramètres photométriques de chaque caméra et d'un changement d'illumination global que nous calculons durant le recalage.

2.2.2 Estimation de fonctions de déformation image en environnement déformable

I34 **Generalized Thin-Plate Spline Warps**
A. Bartoli, M. Perriollat and S. Chambon
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

J11 **Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation**
A. Bartoli
Journal of Mathematical Imaging and Vision, special issue : tribute to Peter Johansen, accepted December 2007

I18 **Direct Estimation of Non-Rigid Registrations**
A. Bartoli and A. Zisserman
BMVC'04 - *British Machine Vision Conf.*, London, UK, September 2004

I40 **Feature-Driven Direct Non-Rigid Image Registration**
V. Gay-Bellile, A. Bartoli and P. Sayd
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007
Version en Français : [N12]

I46 **Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning**
V. Gay-Bellile, A. Bartoli and P. Sayd
ICCV'07 - IEEE *Int'l Conf. on Computer Vision, Rio de Janeiro, Brazil*, October 2007
Version en Français : [N14]
Articles connexes : [I47,N13]

Les deux premiers articles portent sur les fonctions de déformation TPS. Ils permettent d'estimer ces fonctions à partir de correspondances de points. Les trois articles suivants concernent l'estimation d'une fonction de déformation en environnement déformable par des approches basées pixel. Ils permettent de recaler des vidéos montrant une surface déformable, et d'en changer l'apparence dans la vidéo originale ou d'en ré-utiliser les déformations pour une autre apparence.

Le premier article donne une interprétation de la fonction de déformation TPS en terme d'une surface déformable observée par une caméra affine. Il propose trois extensions de cette fonction, en combinant deux options : *(i)* que la scène puisse être une surface lisse rigide et *(ii)* que la caméra qui observe soit décrite par un modèle perspectif. Les fonctions de déformation résultantes sont toutes exprimées à l'aide de notre guidage par primitives (2.4) et sous une forme relevée. En particulier, celles basées sur une caméra affine sont sous la forme affine relevée (2.5), et celles basées sur une caméra perspective sont sous une forme perspective relevée, dont la forme générale est (LP est utilisé pour "Lifted Perspective" en Anglais) :

$$\mathcal{W}_{\text{LP}}\left(\mathbf{q};\tilde{\mathsf{L}}\right) \stackrel{\text{def}}{=} \Psi\left(\tilde{\mathsf{L}}^{\mathsf{T}}\nu(\mathbf{q})\right),$$

où $\tilde{\mathsf{L}}$ est une matrice $(l \times 3)$ contenant par exemple les coordonnées homogènes des l centres de déformation, et Ψ est l'opérateur $\Psi(\tilde{\mathbf{q}}) = \frac{1}{\tilde{q}_3}(\tilde{q}_1 \;\; \tilde{q}_2)$. Lorsque la scène est supposée rigide, les matrices L et $\tilde{\mathsf{L}}$ dépendent de la géométrie épipolaire entre les images sources et cibles.

Le deuxième article porte sur l'estimation des fonctions de déformation sous forme affine relevée avec un lissage externe. Le paramètre de lissage est choisi automatiquement par validation croisée. Cet article montre que la formule usuelle de validation croisée s'étend facilement au cas considéré, et que ces formules sont en général de bonnes approximations du vrai score de validation croisée.

Le troisième article propose une méthode d'estimation des fonctions de déformation RBF avec une procédure d'insertion dynamique de centres de déformation. L'idée est de partir d'un modèle de déformation très simple, par exemple affine. On observe l'image d'erreur après estimation de ce modèle. Les zones avec des valeurs élevées sont interprétées comme des erreurs de recalage, et notamment par le fait que la fonction de déformation n'est pas assez flexible à ces endroits. Nous y insérons donc des centres de déformation, et relançons la procédure de recalage. Nous proposons dans l'article une extension de l'algorithme au cas multi-images.

Le quatrième article apporte deux contributions principales. La première est un ensemble d'outils permettant d'appliquer le principe de composition inverse à des fonctions de déformation qui n'ont pas de structure de groupe. L'idée est basée sur le guidage par primitive, qui permet d'approximer l'inversion et la composition de ces fonctions de manière simple et rapide. La deuxième contribution est une méthode de recalage local basée sur l'apprentissage de la relation entre l'image de différence et les paramètres de mise à jour. Le modèle que nous apprenons est linéaire par morceaux. Il est précis sur les petites déformations et robuste sur les grandes. Un critère de sélection de la partie linéaire la plus pertinente est de même appris.

Le cinquième article est dédié à la gestion des auto-occultations. Lorsqu'une surface se déforme, il est fréquent qu'une partie soit occultée par une autre. Ceci, contrairement aux occultations externes générales, crée des discontinuités dans la fonction de déformation, car les pixels auto-occultés dans l'image cible, mais visibles dans l'image source, doivent être transférés le long de la frontière d'auto-occultation. Nous présentons un module de détection des auto-occultations qui permet, à l'aide d'un terme supplémentaire dans la fonction de coût, de gérer ce type de phénomènes.

2.3 Reconstruction 3D en environnement déformable

Note : La version Anglaise détaillée de cette section se trouve au chapitre 7 dans la partie II de ce document.

Nous étudions le problème du calcul de la structure 3D et de la pose d'un capteur dans un environnement déformable. Dans la première partie, nous présentons cinq articles supposants qu'une caméra unique en mouvement observe la scène. La difficulté est que retrouver des informations 3D est a priori un problème mal posé. La deuxième partie présente quatre articles dédiés au cas où la structure 3D est obtenue à chaque instant par un ensemble de caméras synchronisées ou un capteur de profondeur. La difficulté est de mettre en relation ces structures 3D déformables, et de calculer le mouvement du capteur.

2.3.1 Cas d'une seule caméra

I17 **Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences**
A. Bartoli, E. von Tunzelmann and A. Zisserman
CVPR'04 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Washington, DC, USA, June 2004

I22 **A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery**
A. Bartoli and S. Olsen
WDV'05 - *Workshop on Dynamical Vision* at ICCV'05, Beijing, China, October 2005
Autre version : [N08]

J10 **Implicit Non-Rigid Structure-from-Motion with Priors**
S. Olsen and A. Bartoli
Journal of Mathematical Imaging and Vision, special issue : tribute to Peter Johansen, accepted December 2007
Version antérieure : [I42]

I50 **Coarse-to-Fine Low-Rank Structure-from-Motion**
A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen and P. Sayd
CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

I29 **Image Registration by Combining Thin-Plate Splines with a 3D Morphable Model**
V. Gay-Bellile, M. Perriollat, A. Bartoli and P. Sayd
ICIP'06 - *Int'l Conf. on Image Processing*, Atlanta, GA, USA, October 2006

Les trois premiers articles utilisent le LRSM implicite (2.8), alors que le quatrième article utilise sa forme explicite (2.7). Le cinquième article est basé sur un modèle 3D pré-appris de la forme (2.6).

Le premier article utilise des correspondances de points et de courbes pour estimer le LRSM implicite. L'idée est d'introduire des correspondances de points virtuels le long des courbes, et de chercher directement leurs formes de base implicites. Un ensemble de fonctions de déformation RBF guidées par le LRSM est utilisé, et sert notamment à la vérification du bon recalage au niveau des courbes.

Le deuxième article propose l'utilisation des contraintes de fermeture présentées en §2.1.4 pour l'estimation du LRSM implicite. La prédiction des données est très bonne. Le modèle étant très flexible, il extrapole par contre assez mal.

Le troisième article est une extension du précédent. Il propose deux a priori naturels génériques. Le premier est que la caméra a une trajectoire lisse et que les déformations sont lisses. Le deuxième est que la scène observée est proche d'une surface lisse. Ces deux a priori permettent au LRSM de se généraliser à de nouvelles données.

La quatrième article propose l'approche "coarse-to-fine" pour la reconstruction 3D du LRSM explicite. Les principes de "Deformotion" sont appliqués (Yezzi and Soatto, 2003), et conduisent à l'estimation de la pose et d'une forme moyenne par reconstruction 3D rigide. Des formes de base sont ensuite ajoutées jusqu'à ce que la validation croisée du modèle se détériore. Les deux a priori définis ci-dessus sont utilisés. Les résultats de l'algorithme sont très stables.

Le cinquième article combine un modèle 3D "morphable" avec des fonctions de déformation TPS. Ceci permet un calcul basé pixel. Une approche compositionnelle avec apprentissage est utilisée. Les bases du modèle sont apprises à partir de surfaces générées par la méthode (Salzmann et al., 2007b).

2.3.2 Cas de plusieurs caméras synchronisées et des capteurs de profondeur

I36 **A Quasi-Minimal Model for Paper-Like Surfaces**
M. Perriollat and A. Bartoli
BenCOS'07 - ISPRS *Int'l Workshop "Towards Benmarking Automated Calibration, Orientation, and Surface Reconstruction from Images"* at CVPR'07, Minneapolis, USA, June 2007
Versions antérieures : [I27,N07]
Version en Français : [N11]

I25 **Towards 3D Motion Estimation from Deformable Surfaces**
A. Bartoli
ICRA'06 - IEEE *Int'l Conf. on Robotics and Automation*, Orlando, Florida, USA, May 2006
Version antérieure : [I21]

I48 **Automatic Quasi-Isometric Surface Recovery and Registration from 4D Range Data**
T. Collins, A. Bartoli and R. Fisher
BMVA *Symposium on 3D Video - Analysis, Display and Applications*, London, UK, February 2008

I38 **Joint Reconstruction and Registration of a Deformable Planar Surface Observed by a 3D Sensor**
U. Castellani, V. Gay-Bellile and A. Bartoli
3DIM'07 - *Int'l Conf. on 3D Digital Imaging and Modeling*, Montréal, Québec, Canada, August 2007

Les deux premiers articles utilisent des nuages de points éparses. Les deux suivants supposent qu'une reconstruction dense est disponible.

Le premier article porte sur la reconstruction 3D d'une surface de type feuille de papier, modélisée par une surface développable. Nous proposons une paramétrisation de ces surfaces en termes de règles et d'angles de pliage, ainsi qu'un algorithme de reconstruction 3D. L'idée générale est de mettre en relation automatiquement le contenu de documents écrits et les fonctionnalités informatiques.

Le deuxième article utilise un LRSM explicite en 3D afin de pouvoir calculer la pose d'une paire de caméras synchronisées. Le calcul de ce modèle est un problème non-linéaire, pour lequel nous proposons une solution basée sur des tenseurs d'appariement 3D calibrés. Cet article contraste avec les autres travaux qui calculent le flot de la scène mais non la pose du capteur.

Le troisième article est à propos du recalage d'images de profondeur denses, couplé avec la segmentation automatique de l'objet d'intérêt et la construction de son modèle. L'hypothèse de base est que l'objet d'intérêt se déforme de manière isométrique. Cette hypothèse est approximativement satisfaite par un grand nombre de cas en pratique : le papier, les vêtements ou encore les visages. Notre algorithme est basé sur une segmentation spectrale utilisant des mesures de compatibilité entre correspondances de points. Le calcul est rapide et extrêmement fiable. Une mosaïque de la surface applatie est finalement obtenue. Les données que nous avons utilisées proviennent d'un capteur stéréoscopique trinoculaire.

Le quatrième article porte sur l'estimation d'un modèle de surface et de ses déformations à partir de données de profondeur. Le capteur utilisé ne donne que les points 3D, mais pas d'information d'apparence. La surface est "accrochée" aux données par une détection des bords grâce aux ruptures de discontinuité. La méthode gère les données erronées grâce à l'estimateur robuste X84. Elle est basée sur un algorithme ICP ("Iterated Closest Point" en Anglais) déformable avec transformée en distance.

2.4 Reconstruction 3D en environnement rigide

Note : La version Anglaise détaillée de cette section se trouve au chapitre 8 dans la partie II de ce document.

Cette section porte sur la reconstruction 3D à partir de différents types de primitives. Les contributions sont présentées par type de primitive : points, droites et courbes.

2.4.1 Reconstruction 3D avec des points

I33 **Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion**
J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert and S. Roy
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

I32 **On Constant Focal Length Self-Calibration From Multiple Views**
B. Bocquillon, A. Bartoli, P. Gurdjos and A. Crouzil
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007
Version en Français : [N10]

I23 **Handling Missing Data in the Computation of 3D Affine Transformations**
H. Martinsson, A. Bartoli, F. Gaspard and J.-M. Lavest
EMMCVPR'05 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, Florida, USA, November 2005
Version en Français : [N06]
Version antérieure : [I20]

Le premier article prend en entrée des correspondances de points sur plusieurs images, et calcule une reconstruction 3D par les méthodes de factorisation de matrice présentées en §2.1.4. Les idées directrices de ces méthodes ont été introduites dans cet article. La spécificité est liée à la partie translationnelle affine, dans le cas de caméras affines, et aux profondeurs projectives, dans le cas de caméras perspectives. Le calcul est très rapide, précis et robuste : une classification des points image erronés est obtenue. Les méthodes de moindres carrés non-linéaires convergent en quelques itérations vers le minimum global lorsqu'elles sont initialisées par notre estimation, et ce sur plusieurs jeux de données réelles et simulées.

Le deuxième article étudie le problème du calibrage en ligne d'une caméra avec pour seule inconnue sa distance focale fixe. Nous donnons l'ensemble des séquences de mouvements critiques qui ne permettent pas le calibrage en ligne. Un algorithme de calibrage stratifié est proposé. Cet algorithme est basé sur l'optimisation par analyse par intervalles, ce qui permet de garantir que le minimum global de la fonction de coût non-linéaire est obtenu.

Le troisième article propose une méthode de recalage de deux reconstructions 3D obtenues à partir de caméras affines. C'est un des composants essentiels pour les méthodes de reconstruction 3D hiérarchiques où

des modèles 3D partiels doivent être fusionnés. Notre contribution principale est un algorithme qui permet de minimiser une bonne approximation de l'erreur de reprojection dans toutes les images des deux ensembles de caméras par une simple SVD. Les données manquantes sont gérées par EM ("Expectation-Maximization" en Anglais).

2.4.2 Reconstruction 3D avec des droites

I16 **A Framework for Pencil-of-Points Structure-From-Motion**
A. Bartoli, M. Coquerelle and P. Sturm
ECCV'04 - *European Conf. on Computer Vision*, Prague, Czech Republic, May 2004

J09 **Triangulation for Points on Lines**
A. Bartoli and J.-T. Lapresté
Image and Vision Computing, Vol. 26, No. 2, p. 315-324, February 2008
Version antérieure : [I26]

I35 **Kinematics From Lines in a Single Rolling Shutter Image**
O. Ait-Aider, A. Bartoli and N. Andreff
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

Les deux premiers articles sont très liés car ils utilisent tous deux des points sur des droites.

Le premier article introduit un nouveau type de primitive composite que nous appelons pinceaux de points (POP).[13] Nous montrons que ces POPs possèdent de bonnes propriétés en terme de répétabilité de détection. Nous étudions toutes les étapes pour la reconstruction 3D à partir de POPs : détection, mise en correspondance, estimation des tenseurs d'appariement et triangulation. Notons que trois POPs suffisent à définir la géométrie épipolaire, dont le calcul par RANSAC devient alors extrêmement rapide.

Le deuxième article étudie le problème de la triangulation d'un point sur une droite. C'est une des étapes pour l'inférence de la structure 3D à partir de POPs. Nous proposons un algorithme polynomial qui garantit que l'erreur de reprojection est minimisée. Il se trouve que le degré du polynôme à résoudre dépend linéairement du nombre d'images. Ceci permet de minimiser l'erreur de reprojection sur des centaines d'images en une fraction de seconde.

Le troisième article est basé sur les caméras de type "rolling shutter" pour le calcul de la pose et de la cinématique instantanée entre un objet et la caméra. Ces caméras acquièrent les lignes de l'image séquentiellement. Les droites de l'espace en mouvement sont donc projetées en des courbes. Nous instancions des points le long des droites 3D et minimisons la distance entre leur reprojection et la courbe correspondante, ainsi que sur les paramètres du modèle de caméra, qui inclut la pose et la cinématique. La minimisation est effectuée par l'algorithme de Levenberg-Marquardt. Nous utilisons le fait que la matrice Hessienne approchée a une structure par bloc similaire à celle habituellement obtenue en ajustement de faisceaux.

2.4.3 Reconstruction 3D avec des courbes pour le contrôle qualité

I30 **Reconstruction of 3D Curves for Quality Control**
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007
Version en Français : [N09]

I37 **Energy-Based Reconstruction of 3D Curves for Quality Control**
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
EMMCVPR'07 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EZhou, Hubei, China, August 2007
Version connexe : [I45]

[13] "Pencil-of-Points" en Anglais.

Nous présentons deux articles traitant de la reconstruction 3D de courbes pour le contrôle qualité d'objets manufacturés. L'idée est de comparer les courbes reconstruites à celles du modèle CAD afin de découvrir les éventuels défauts de fabrication. Nous utilisons des courbes NURBS ("Non-Uniform Rational B-Splines" en Anglais) car leur utilisation est très répandue dans le domaine de la conception graphique, et nous les rencontrons dans les modèles CAD. Elles ont l'avantage du contrôle local, la possibilité d'insérer facilement de nouveaux points de contrôle et sont projectivement covariantes. Nous procédons en itérant deux étapes. La première rafine les paramètres de la courbe à reconstruire, initialisée par le modèle CAD. La deuxième insère des points de contrôle aux endroits où la courbe se reprojette mal dans les images. Nos deux articles utilisent des termes de donnée basés primitive et pixel respectivement : une distance géométrique entre des points et la courbe, et le gradient image le long de la courbe. Nos expérimentations sur données simulées et réelles montrent que la méthode basée pixel est en général la plus précise.

2.5 Autres travaux

Note : La version Anglaise détaillée de cette section se trouve au chapitre 9 dans la partie II de ce document.

Les travaux décrits dans cette section sont liés par plusieurs aspects au reste de ce document. La première partie comprend deux articles sur les AAMs. La deuxième partie concerne les critères PRESS et LOOCV.

2.5.1 Les modèles d'apparence actifs

I41 **Segmented AAMs Improve Person-Independent Face Fitting**
J. Peyras, A. Bartoli, H. Mercier and P. Dalle
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007

I49 **Light-Invariant Fitting of Active Appearance Models**
D. Pizarro, J. Peyras and A. Bartoli
CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

Le premier article adresse le problème de la construction et de l'ajustement d'un AAM à des personnes inconnues (*i.e.* qui ne sont pas dans les images d'apprentissage). C'est un problème difficile car la variabilité identitaire est évident très grande. Nous proposons les AAMs segmentés multi-niveaux. L'idée est de décoreller les différentes parties du visage : les sourcils, les yeux, le nez et la bouche. Ceci améliore grandement la capacité de l'AAM à générer de nouvelles identités. L'ajustement de l'AAM à une image est assuré par la partie multi-niveaux du modèle. Un AAM global est tout d'abord ajusté. Le résultat n'est pas précis, mais plus robuste qu'avec des modèles locaux. Il permet donc d'initialiser des modèles intermédiaires, qui eux-même initialiseront les modèles locaux.

Le deuxième article traite le problème de l'ajustement d'un AAM à une image en présence de variations d'illumination. Ces dernières entraînent la création d'ombres portées, ce qui empêche le bon fonctionnement de l'AAM. Les méthodes habituelles essayent de modéliser ces ombres portées. Ceci alourdi l'AAM, et ne fonctionne pas lorsque les ombres sont portées par un élément externe, et non par le visage lui-même, car la variabilité est trop élevée. Nous proposons d'utiliser les espaces 1D invariants que nous avons déjà utilisés en §2.2.1 pour le recalage d'images. L'idée est de projeter dans cet espace l'image générée par l'AAM et l'image requête afin d'y mesurer une erreur d'ajustement non perturbée par les ombrages. Les paramètres photométriques des caméras sont estimés lors de l'ajustement de l'AAM.

2.5.2 La "Prediction Sum of Squares statistic" et la validation croisée

On Computing the Prediction Sum of Squares Statistic in Linear Least Squares Problems with Multiple Parameter or Measurement Sets
A. Bartoli
Submitted to IEEE *Transactions on Neural Networks*, December 2007

N15 **Reconstruction de surface par validation croisée**
F. Brunet, A. Bartoli, R. Malgouyres et N. Navab
ROADEF'08 - *Journées de recherche opérationnelle et d'aide à la décision*, Clermont-Ferrand, France, février 2008

Le premier article étudie des formules non itératives permettant de calculer le PRESS et le LOOCV. Ces formules ont été établies pour un certain nombres de cas. Leur application à l'estimation des fonctions de déformation rigides décrites en §2.2.2 n'est cependant pas directe, car il s'agit de problèmes de moindres carrés linéaires non standards. Nous étudions les combinaisons des cas où des paramètres sont liés et où des mesures sont liées. Pour chaque cas, nous donnons la formule permettant de calculer le PRESS et le LOOCV de manière non itérative.

Le deuxième article porte sur la reconstruction d'une surface à partir de données 2,5D denses. L'idée est de sélectionner le paramètre de lissage par validation croisée. Le modèle de surface exprime l'élévation en fonction des coordonnées dans une image par une déformation de forme libre.

CHAPITRE

3

PERSPECTIVES

Formuler des perspectives de recherche scientifique pertinentes est important. Elles doivent pour moi être formulées avec toute la passion et la curiosité requises par le métier de chercheur, et trouver leur cohérence avec le système et les structures de recherche à différentes échelles.

La première section donne mes perspectives scientifiques. Elles sont articulées avec les sujets de recherche des post-doctorants et doctorants que j'encadre ou co-encadre actuellement. La deuxième section montre comment mes recherches s'intègrent à la communauté et aux structures scientifiques locales, nationales et internationales, d'un point de vue recherche et applicatif. La troisième section est centrée sur mes perspectives de transfert technologique. La première section est reprise au chapitre 10 dans la partie II de ce document.

3.1 Perspectives scientifiques

Les perspectives données ci-dessous sont situées par rapport à mes résultats de recherche sur la période 2004 – 2007. Mes contributions et résultats concernent le recalage d'images et la reconstruction 3D de la structure et de la caméra en environnement rigide et déformable. Mes perspectives sur le court et le long terme concernant ces problématiques sont détaillées en §§3.1.2 et 3.1.3. Elles sont pour la plupart liées au problème général du choix automatique de la complexité d'un modèle, décrit en §3.1.4. Ceci est une problématique que j'entrevois comme une des lignes directrices pour mes futurs travaux de recherche en §3.1.1.

Trois des doctorants que je co-encadre ont commencé la rédaction de leur thèse. Nous avons planifié des soutenances autour de septembre 2008. Mathieu Perriollat a travaillé sur la modélisation, la paramétrisation et la reconstruction 3D des surfaces de type papier. Nous pensons utiliser ces résultats comme point de départ pour la conception d'un prototype à destination de l'industrie, comme décrit en §3.1.5. Vincent Gay-Bellile a principalement travaillé sur le recalage déformable d'images. Mes perspectives sur ce sujet sont données ci-dessous. Hanna Martinsson a travaillé sur la reconstruction 3D rigide pour le contrôle qualité d'objets manufacturés. Nous avons obtenu des résultats prometteurs par reconstruction de courbes 3D. J'ai par ailleurs encadré des doctorants d'autres universités en visite au laboratoire, comme par exemple Jean-Philippe Tardif qui était alors inscrit à l'Université de Montréal. Des perspectives sur notre travail sont données en §3.1.6.

3.1.1 Problématiques transverses

Le bilan de mes travaux, réalisé lors de la rédaction de ce document, m'amène à afficher une nouvelle problématique qui leur est transverse : la sélection automatique de la complexité d'un modèle. Cette problématique complexe s'est en effet imposée comme étant commune à plusieurs problèmes de recalage d'images et de reconstruction 3D rigide et déformable. Je désire l'inscrire comme une des problématiques transverses que j'aborde, et étendre les études conduites pour la vision par ordinateur à un cadre plus général. Voici l'énoncé de cette nouvelle problématique, la description des deux premières (*mise en relation d'images* et *reconstruction 3D*) restant identique à celle donnée en §1.3.3 :

3. **Sélection de la complexité d'un modèle.** Etant donné un modèle "flexible" (dans le sens où le nombre de paramètres est variable et où certains paramètres peuvent être contraints par des pénalités de lissage), comment déterminer la complexité "optimale" du modèle ?

Cette nouvelle problématique est particuliérement importante pour la vision en environnement déformable, mais est aussi reliée à des problèmes en environnement rigide. De nombreux détails sont donnés ci-dessous.

De manière plus générale, il y a de nombreux algorithmes en apprentissage artificiel utiles pour résoudre des problèmes de vision par ordinateur. Citons par exemple la réduction de dimension, l'ACP à noyau, l'ajustement de variétés et la sélection d'indices.

3.1.2 Reconstruction 3D rigide avec calibrage en ligne et lissage de la trajectoire

La reconstruction 3D par vision est depuis récemment embarquée dans des modules de localisation comme un complément au GPS (Global Positioning System). Le calibrage en ligne de la caméra est important car ses paramètres peuvent changer. La communauté scientifique s'accorde sur le fait que la plupart des résultats scientifiques sur ce sujet ont été trouvés. Il persistent cependant des problèmes perturbant la fiabilité des systèmes, et notamment la détection des situations instables ou dégénérées. Une façon d'améliorer la précision et la fiabilité est d'incorporer des a priori sur le problème, tel le fait que la trajectoire de la caméra soit continue, voire lisse. L'idée est de combiner l'erreur de reprojection classique, c'est-à-dire le terme d'attache aux données, avec une pénalité de lissage. Ceci n'est pas nouveau. Il faut cependant prendre en compte que le paramètre de lissage, qui contrôle l'influence de la pénalité de lissage, doit être automatiquement estimé. La pénalité de lissage doit typiquement être renforcée pour les configurations instables. J'identifie deux axes de recherche importants :

- ▷ **Estimation du paramètre de lissage en temps-réel.** Le paramètre de lissage doit être spécifiquement adapté à chaque image du flux vidéo. Son calcul doit être rapide, permettant au système de traiter les images à la volée, et de donner par exemple un retour immédiat à un véhicule autonome.
- ▷ **Utilisation conjointe de plusieurs pénalités de lissage.** Comme mentionné ci-dessus, la trajectoire de la caméra est au moins continue, et parfois lisse. Utiliser réellement ces a priori sur le problème nécessite l'utilisation de plusieurs pénalités de lissage avec des paramètres de lissage individuels et adaptatifs. Ceci ajoute à la difficulté d'un calcul en temps-réel.

J'ai commencé à travailler sur le calcul d'un score de validation croisée en temps réel pour la reconstruction 3D séquentielle avec Michela Farenzena, post-doctorante au laboratoire, et Youcef Mezouar de l'équipe ROSACE.

3.1.3 Recalage déformable d'images et reconstruction 3D monoculaire

En dépit d'avancées récentes significatives, la reconstruction 3D monoculaire en environnement déformable reste un problème ouvert. Le but est de concevoir un système aux performances comparables à celles des systèmes de reconstruction 3D rigide. La différence principale réside dans le fait que l'utilisation d'a priori sur le problème est incontournable afin de rendre ce dernier bien posé. Les axes de recherche suivants me semblent pertinents :

- ▷ **Recalage d'images.** La mise en correspondance d'images d'un environnement déformable est en général difficile. Ceci est dû au manque de contraintes géométriques telle que la géométrie épipolaire. Ceci est bien sûr fortement lié à la représentation utilisée pour la forme 3D. Par exemple, il est communement

admis qu'un ensemble éparse de correspondances de points permet d'estimer la pose de la caméra de manière robuste. Nos contributions sur les tenseurs d'appariement de faible rang montre qu'il est possible d'améliorer le suivi de points dans des environnements non structurés. Dans le cas d'une surface continue telle une feuille de papier, il est cependant possible de mieux exploiter les contraintes photométriques. Ceci permettrait de bénéficier des avancées récentes en suivi développées au CVLAB à l'EPFL, et des méthodes de recalage précisent que nous avons proposées.

- ▷ **A priori génériques et spécifiques.** La reconstruction 3D monoculaire déformable n'est possible que si des a priori sur la structure 3D et / ou la caméra sont utilisés. Des a priori importants sont donnés par ce que j'appelle les guides statistiques multilinéaires, et en particulier le modèle de faible rang non pré-appris. Il est montré en §7.1.4 que l'utilisation d'a priori génériques comme le lissage de la trajectoire de la caméra améliore significativement la précision et la stabilité de la reconstruction 3D. Les a priori forment un continuum : il n'y en a pas de totalement générique ou spécifique. Je pense que trouver de nouveaux a priori est un sujet de recherche encore ouvert, qui n'a de limite que l'imagination.
- ▷ **Traitement séquentiel, complétion et mise à jour du modèle 3D.** Les modèles déformables sont appris – en ligne ou hors ligne – à partir d'images. Un exemple est l'estimation des formes de base du modèle de faible rang. Le problème a été traité en supposant toutes les images disponibles. Il serait cependant très utile de pouvoir mettre à jour un modèle déformable de manière séquentiel à mesure que les images arrivent.

Samir Khoualed a démarré sa thèse sous ma direction en octobre 2007. Il travaille sur les fonctions d'observation basées sur des points clefs. Nous prévoyons de travailler sur les modèles statistiques de forme. Le sujet de thèse de Dawei Liu, que je co-encadre depuis septembre 2007, concerne l'utilisation de modèles issus de la mécanique des milieux continus pour la formulation d'a priori en reconstruction 3D de surfaces déformables. Nous travaillons avec l'équipe M&M du LaMI. Finalement, je co-encadre Pauline Julian qui fait une thèse sur un contrat CIFRE avec l'entreprise FittingBox basée à Toulouse. Elle travaille sur le suivi de visage et la reconstruction 3D avec des AAMs.

3.1.4 Méthodes générales pour l'ajustement automatique de la complexité d'un modèle

Comme mentionné ci-dessus, la sélection automatique de la complexité d'un modèle est une étape clef et critique pour de nombreux algorithmes. Il existe plusieurs approches, et de nombreuses voies de recherche sont encore ouvertes, notamment :

- ▷ **Des critères rapides, stables et garantis.** Il y a besoin de critères de prédictivité d'un modèle rapides à calculer, numériquement stables, et pour lesquels une méthode garantie trouve la meilleure solution. Prenons l'exemple de la validation croisée. J'ai proposé des formules non itératives pour estimer des modèles de déformation image. Il n'y a cependant aucune garantie qu'en pratique le minimum sélectionné corresponde à la meilleure solution, ni que le critère de validation croisé ait un seul minimum. Les algorithmes de recalage et de reconstruction 3D mentionnés ci-dessus nécessitent un critère peu coûteux ainsi qu'un algorithme dont le résultat est garanti. Je pense que l'approche qui consiste à ajouter des centres de déformation à un modèle déformable tout en mesurant le "Prediction Sum of Squares statistic" suivie en §9.2.1 est prometteuse.
- ▷ **Combiner plusieurs termes de lissage.** Les fonctions de coût sont souvent basées sur un terme de donnée et un terme de lissage. Chaque terme de lissage est lié à un a priori sur le problème. Il est donc désirable de pouvoir combiner plusieurs termes de lissage, afin de modéliser des a priori complexes et détaillés sur le problème. Cela pose le problème du calcul des paramètres de lissage, car les critères existants ne se généralisent pas tous à plusieurs termes de lissage (par exemple, la validation croisée peut se généraliser), et le coût croit en complexité. L'utilisation de plusieurs termes de lissage pourrait par exemple ouvrir des perspectives en reconnaissance d'objets, par apprentissage de termes de lissage spécifiques à des classes d'objets.

- ▷ **Nombre de paramètres libres et nombre effectif de paramètres.** Ces notions sont liées aux deux méthodes de contrôle de la complexité d'un modèle : l'ajout et la suppression directs de paramètres et l'utilisation de lissage pour les contraindre. J'ai essayé ces deux méthodes, en utilisant respectivement la validation croisée et la "Prediction Sum of Squares statistic". Je pense qu'elles ont des propriétés différentes et complémentaires. Leur combinaison devrait donner des résultats intéressants.
- ▷ **Sélection de complexité et robustesse.** Les critères de prédictivité que j'ai utilisés ne sont pas robustes, dans le sens de la prise en compte de données erronées. Ceci n'est pas problématique pour les problèmes "très contraints" tels que la reconstruction 3D rigide, pour lesquels RANSAC peut être utilisé initialement afin de rejeter les fausses correspondences. C'est en revanche un problème ouvert pour les environnements déformables, beaucoup moins contraints. Un sujet de recherche intéressant est la combinaison des critères de prédictivité avec des méthodes robustes telles RANSAC.

Une autre possibilité de recherche est l'utilisation de critères de prédictivité avec des méthodes basées pixel. Cela soulève le problème du "test sur les données d'apprentissage", car les données sont très denses et corrélées.

Florent Brunet a commencé sa thèse en octobre 2007. Je suis un de ses co-encadrants. Son sujet de recherche est la reconstruction dense de surface 2,5D. Nous travaillons actuellement sur un critère de prédictivité basé sur la "$\mathcal{L}$-curve" (Lawson and Hanson, 1974) ayant les propriétés mentionnées ci-dessus.

3.1.5 Reconstruction 3D de papier

Nous avons contribué à différents aspects nécessaires à la reconstruction 3D de feuilles de papier à partir d'images. Ceci inclut la modélisation mathématique des surfaces déformables, leur paramétrisation, et des algorithmes d'estimation. Nous pensons qu'une large part des algorithmes de reconstruction 3D est maintenant mature. Nous envisageons un transfert technologique vers l'industrie. Ces résultats ont été obtenus avec Mathieu Perriollat, un doctorant que je co-encadre.

3.1.6 Factorisation d'une matrice avec données manquantes et erronées

Les algorithmes de factorisation de matrice que nous avons initialement proposés pour la reconstruction 3D ont été généralisés[1] en §2.1.4. Ils fournissent des méthodes de moindres carrés linéaires permettant de trouver une solution qui en pratique est proche de la solution optimale du problème non-linéaire. Ils sont efficaces au sens vitesse de calcul et ont atteint un niveau de développement mature. Un de nos buts est l'écriture et le partage d'une librairie implémentant ces algorithmes. Ces résultats ont été obtenus en collaboration avec plusieurs chercheurs, et en particulier Jean-Philippe Tardif qui est maintenant en post-doc à l'Université de Pennsylvanie.

3.2 Intégration à la communauté scientifique et déclinaisons applicatives

Je décris dans cette section comment mes travaux passés et mes perspectives s'inscrivent à différents niveaux des structures et communautés scientifiques, au travers de leurs aspects fondamentaux et de certaines de leurs déclinaisons applicatives.

3.2.1 Structures et partenaires locaux

Il y a trois éléments clefs évoqués ci-dessous : l'équipe de recherche ComSee, le groupe GRAVIR du LASMEA, et la Fédération de Recherche TIMS. Ils représentent les structures de recherche dans lesquels mes travaux s'inscrivent principalement au niveau de la place Clermontoise.

L'équipe de recherche ComSee est décrite en §1.3. L'axe de recherche "vision en environnement déformable" est clairement moteur, que ce soit au niveau des résultats scientifiques, des publications, des contacts et du rayonnement. Il est en interaction importante avec les deux autres axes de l'équipe, de par les problématiques

[1] Un article dont le premier auteur est Jean-Philippe Tardif a été soumis à un congrès.

de base communes qui y sont abordées. Il doit être renforcé au niveau personnel, car je suis le seul permanent impliqué. Le recrutement d'un jeune enseignant-chercheur sur cet axe est la priorité de l'équipe, et de GRAVIR. Je désire développer cet axe en intensifiant les collaborations avec les autres équipes au niveau de GRAVIR.

Cet axe trouve toute sa place et sa cohérence au sein de TIMS et du projet CPER Innov@Pôle. Nous participons activement au projet "Modèles et Logiciels pour la Santé, le Vivant et le Physique" (MLSVP). La thèse de Florent Brunet est notamment co-encadrée par moi-même, Nassir Navab de la TUM et Rémy Malgouyres du LAIC. Les aspects mise en relation d'images, reconstruction 3D et surfaces déformables sont des ingrédients essentiels de ce projet. Nous collaborons avec Laurent Sarry de l'ERIM sur un projet de reconstruction de "stents" coronaires par tomographie optique cohérente (OCT). Nous participons au projet "Machines et Mécanismes Innovants" (M2I) à travers une collaboration avec Michel Grédiac du LaMI, sur le thème de la caractérisation de matériaux par vision. Finalement, nos perspectives sont très orientées sur l'apprentissage articifiel, et rentrent tout à fait dans le cadre d'un groupe de travail récemment créé sur ce thème au sein de TIMS. TIMS est une force, car elle met en relation les gens et les compétences au sein des différents laboratoires Clermontois. Il est essentiel d'exploiter cette structure pour la cohérence des recherches, tout en favorisant les contacts directs avec les structures et partenaires au niveau national et international.

Finalement, je voudrais souligner mon investissement dans d'autres tâches au niveau local : enseignement (notamment au sein du Master VIRO), mise en place des séminaires de GRAVIR, participation à diverses commissions (recrutement, relations internationales), *etc.*

3.2.2 Communauté et structures nationales

Mes recherches s'inscrivent pleinement au niveau du Groupement de Recherche CNRS Information, Signal, Images et Vision (GDR ISIS), dans le cadre duquel j'ai donné deux présentations invités en 2006. Je publie régulièrement aux congrès RFIA et ORASIS. Mes recherches sont bien implantées au sein de la vision par ordinateur, mais ont aussi un caractère original. Par exemple, peu de groupes Français utilisent le LRSM pour la reconstruction 3D monoculaire en environnement déformable.

Un indicateur est celui de ma participation à des projets nationaux : HFIBMR ("High Fidelity Image Based Modeling and Rendering") est un projet financé par l'ANR avec pour partenaires l'équipe WILLOW (ENPC / ENS / INRIA, Paris), le LASMEA et l'équipe ARTIS (INRIA Rhône-Alpes), PMoCap ("Computer Vision Based Motion Capture for Paper") est un projet commun avec l'INRIA Rhône-Alpes financé par le GDR ISIS et STANDS-MSG ("Spatio-Temporal Analysis of Deformation Structures in MSG Images") est financé par l'ANR et commun à différents instituts dont le laboratoire COSTEL à Rennes. Je collabore par ailleurs de manière significative avec l'IRIT de Toulouse. J'ai donné deux fois trois heures de cours de niveau doctoral au LIRMM à Montpellier.

Je désire maintenir et amplifier ce tissu de relations, tout en l'ouvrant à des laboratoires étudiant des domaines connexes et reliés tel l'apprentissage artificiel et les neurosciences. Mes perspectives de recherche s'inscrivent principalement en vision par ordinateur et apprentissage artificiel.

3.2.3 Communauté internationale

La communauté internationale de vision par ordinateur est très active. Plusieurs groupes effectuent des recherches proches des miennes. Je suis présent dans cette communauté au travers de nombreux contacts et publications dans les revues et congrès de reférence (respectivement IJCV, PAMI, CVIU, JMIV et ICCV, CVPR, ECCV, BMVC). Je participe et ai participé à l'organisation d'ateliers associés à des congrès comme CVPR et BMVC, et à un tutoriel associé à ISMAR'07. Je suis dans les comités de programme de la plupart de ces congrès.

Mon statut de "Visiting Professor" dans l'Image Group du DIKU me donne l'occasion de visiter très régulièrement cet institut, d'y enseigner, et de participer aux manifestations scientifiques qui s'y déroulent. Je collabore avec plusieurs personnalités et équipes importantes du domaine.

Mes perspectives de recherche permettent une continuation naturelle de ces relations. Tout comme au niveau national, je désire leur donner une teinte imagerie médicale et apprentissage artificiel. Cette dernière correspond aux inclinaisons récentes de la communauté scientifique.

3.3 Transfert technologique

Je n'ai jusqu'à présent que peu touché au transfert technlogique et à la valorisation industielle de mes travaux. Je fais des consultances auprès de l'entreprise FittingBox de Toulouse, en tant que membre d'un comité d'experts scientifiques (avec Pierre Gurdjos de l'IRIT et Peter Sturm de l'INRIA) mandaté par l'entreprise. Le transfert technologique est une de nos missions en tant que chercheurs et une des finalités de nos recherches. Il est donc important d'y consacrer du temps et de l'énergie lorsque nous sentons que nos recherches arrivent à un point qui y est favorable. Je vois aujourd'hui deux possibilités de transfert autour de mes travaux récents. La première, à court ou moyen terme concerne la reconstruction 3D de papier, comme décrit ci-dessus en §3.1.5. La deuxième, à plus long terme, concerne la caractérisation de matériaux par vision, résultant de ma collaboration au sein de TIMS avec le LaMI.

Part II

Research Results 2004 – 2007

CHAPTER

4

INTRODUCTION

This is the scientific part of my thesis. My contributions fall in the field of computer vision. There are recent textbooks and collections on this field: (Faugeras et al., 2001; Hartley and Zisserman, 2003) focus on the geometry on multiple images while (Forsyth and Ponce, 2003; Paragios et al., 2005) give a broader view.

On the one hand, research in computer vision is motivated to a large extent by the ubiquitous presence of images in modern societies. This is due to the rapid development of cheap computers and imaging sensors. Indeed, mass-market point-and-shoot cameras and webcams provide high-quality still pictures and movies. These sensors are small, can be easily embedded and are not invasive. They highly demand robust vision algorithms and software. Most of the particular computer vision problems I have contributed to have important potential applications. Structure-from-Motion for rigid and deformable environments can be applied in architecture to reconstruct buildings, in the film industry for special effects and in robotics for localization purposes, to name just but a few. Image registration has applications in, for instance, augmented reality for surface retexturing and augmentation, and medical image analysis for multimodal image fusion.

On the other hand, research in computer vision is motivated by intellectual curiosity, and the excitement related to the problem of artificial perception and reasoning. Computer vision is an advanced research topic, tighly bound to artificial intelligence and Machine Learning. This is reflected by some of the current trends in the community, where Machine Learning techniques are getting used more and more for tasks such as visual tracking and recognition. Machine Learning is an active area of research. Some useful textbooks include (Bishop, 1995; Hastie et al., 2001).

A digital image is given by a photometric sensor (*i.e.* a camera). It is the result of the light interacting with the scene structure, which can be rigid or deforming. Through the course of making my contributions, I have tackled two main research areas:[1]

1. **Image matching.** Given several images of the same scene (or several images showing semantically equivalent contents, such as faces), how to match the pixels of these images or geometric features such as keypoints?

[1] These are the same as those of the ComSee research team that I am co-leading, as described in §1.3.3.

2. **3D reconstruction.** Given one or more images of the same scene, how to represent and compute the 3D scene structure and the camera pose?

The physical image formation process is highly complex. This raises several questions:

▷ **Explicit modeling versus invariance.** It would be hard to explicitly model and reconstuct everything, or reconstruct the plenoptic function.[2] Choices have therefore to be made on whether a phenomenon should be explicitly modeled or whether the observation function should be made invariant to its effects. Unmodeled phenomena cause outliers, which can be rejected using robust estimation methods. An example of this is the one of lighting in image registration. Explicitly modeling the effect of global ambient lighting can in general be easily done using for instance a gain and a bias on the pixel colors, as in §6.1.2. However, explicitly modeling complicated lighting changes is much more complicated, since this entails us modeling the scene structure, its BRDF[3] and the light source positions. This is a case where resorting to invariance makes life easier, as shown in §§6.1.3 and 9.1.2.

The choices I have made generally follow these rules: global light changes are explicitly modeled, while complex ones are dealt with by invariance. The unknown 3D scene structure is either explicitly modeled or contained in an image level warp. In the former case, a camera is usually modeled as well. I have sometimes used what I call a statistical driver such as a 3D Morphable Model so as to constrain the deformation centres of a warp. The model components are obviously strongly related to the nature of the cost function that is being used; feature-based or pixel-based.

▷ **Genericity, prior knowledge and complexity tuning.** It is universally agreed that the more prior knowledge the better the solution to a problem. It is challenging to try to use prior knowledge which is as generic as possible, in the sense of not being specific to a particular object-class. A typical example of such generic priors are spatial and temporal smoothnesses. Using such priors raises the problem of tuning the extent to which they influence the estimation. This is modeled by *smoothing parameters*. Emphasizing the priors too much makes the estimate too smooth, increasing its bias. On the contrary, a low weight on the priors increases the variance of the estimate. This is a typical Machine Learning problem. It arises in deformable image warp estimation. I have used tools such as Cross-Validation to estimate optimal smoothing parameters, in particular for the warp estimation problem in §6.2.2. The idea is to minimize a generalization error, describing how well the warp predicts new data. The reader is referred to (Poggio et al., 2004) for more details on generalization and predictivity in Machine Learning.

There are several ways of measuring the predictivity of a model. I have made an extensive use of the Prediction Sum of Squares statistic (Allen, 1974) and Cross-Validation (Wahba and Wold, 1975). Generic deformable models typically are empirical and do not allow one to formulate a parametric distribution function of the residuals, which rules out most model selection techniques.

There are at least two ways of tuning the complexity of a model. The first one is to add or remove pieces of the model. This changes the *number of free parameters*. I used this approach in §9.2.1 to determine the number of centres of deformation for a Thin-Plate Spline warp through the Prediction Sum of Squares statistic. The other way of tuning the complexity of a model is to change the smoothing parameter when estimating the model parameters. This changes the so-called *effective number of parameters*, as defined by (MacKay, 1992; Moody, 1992). This is the approach I have taken when computing the smoothing parameter with Cross-Validation in §6.2.2.

4.1 Organization of this Part

Over the past four year, I have tackled various topics, ranging from classical Structure-from-Motion (SfM) to image warp estimation. The solutions I have proposed are based on models and computational tools that I

[2] The 7D plenoptic function characterizes the light rays that can be observed at any wavelength, from any position and orientation, and at any time (Adelson and Bergen, 1991).

[3] Bidirectional Reflectance Distribution Function – it describes how the surface reflects the incoming light.

sometimes improved and can often be applied to several different problems. There are thus at least two ways of looking at my contributions: either by the technical aspects (*e.g.* matrix factorization) or by the goals (*e.g.* image registration). I have chosen to organize the rest of this thesis so that both ways are reflected. The internal organization of chapter 5 follows the former while chapters 6, 7 and 8 follow the latter:

- ▷ Chapter 5 gives an overview of the models and computational methods I have used and contributed towards. It brings most of the basic elements required to understand the content of the remaining chapters and brings a general view of my key contributions.
- ▷ Chapter 6 covers my contributions on image registration. This includes the modeling of the image photometry and the estimation of rigid and deformable image warps.
- ▷ Chapter 7 covers my contributions on deformable SfM. The first section is on the case of monocular image data. The image registration methods proposed in chapter 6 typically are used to provide input data, *i.e.* image correspondences. The second section is on the case of range data.
- ▷ Chapter 8 covers my contributions on rigid SfM. It is organized by the type of features that are used. These have specifically included points, lines and curves. It has a strong visual geometry flavor.
- ▷ Chapter 9 covers our contributions on other topics, such as the fitting of Active Appeance Models.
- ▷ Chapter 10 concludes and gives directions for future work.

The references using the formatting such as in [I50] (*i.e.* with square braces) refer to my personnal bibliography given in the first part of this document in §1.12.

4.2 Notation and Some Mathematical Tools

The notation followed in this thesis is first reviewed. Some papers included in the companion document might not follow the same conventions. I do not explicitly make a distinction between geometric entities and their coordinates in some coordinates frame. For instance, $\mathbf{q}$ is a point, and is also a 2-vector, giving its coordinates. The coordinate frame is often obvious from the context. Column vectors are universally adopted. Row vectors are obtained using the transpose, as in $\mathbf{q}^\mathsf{T}$. Scalars are in italics or greek letters (*e.g.* j, λ), vectors are always in bold font (*e.g.* $\mathbf{q}$, $\mathbf{K}_j$) and matrices in sans-serif and calligraphic fonts (*e.g.* P, K_λ, $\mathcal{U}$). Homogeneous coordinates are represented by a tilde on the letter (*e.g.* $\tilde{\mathbf{q}}$). The 2D 'de-homogenization' function Ψ is defined by $\Psi(\tilde{\mathbf{q}}) = \frac{1}{\tilde{q}_3}(\tilde{q}_1 \;\; \tilde{q}_2)^\mathsf{T}$. Image warps are written with calligraphics fonts (*e.g.* $\mathcal{W}$). The size of a vector or matrix is written as in (3×4), and may be given as a subscript as in $\mathsf{P}_{(3\times4)}$. The identity, zero and 'all-one' matrices are written I, $\mathsf{0}$ and $\mathsf{1}$, while the zero and 'all-one' vectors are written $\mathbf{0}$ and $\mathbf{1}$. The diag operator is similar to the one in MATLAB (*i.e.* it both extracts a matrix diagonal and constructs a diagonal matrix from a vector). $\mathrm{s.t.}$ is used to abbreviate 'such that' in the equations.

The Hadamard element-wise product of equal size matrices is written $\odot$. The vector $\mathcal{L}_2$ norm and the matrix Frobenius norm are as follows:

$$\|\mathbf{x}\|_2 \stackrel{\text{def}}{=} \sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}} \qquad \text{and} \qquad \|\mathsf{A}\|_\mathcal{F} \stackrel{\text{def}}{=} \sqrt{\mathrm{tr}(\mathsf{A}^\mathsf{T}\mathsf{A})},$$

where tr is the matrix trace operator. The Euclidean distance between two points $\mathbf{q}$ and $\mathbf{q}'$ is $d(\mathbf{q},\mathbf{q}') = \|\mathbf{q}-\mathbf{q}'\|_2$. Integer and real numbers are respectively written $\mathbb{N}$ and $\mathbb{R}$ with the dimension c indicated as in $\mathbb{R}^c$. The projective space of dimension c is written $\mathbb{P}^c$. The solution to an optimization problem is written with a star, such as $\mathbf{u}^\star$.

As for image registration, the source and target images are respectively written $\mathcal{S}$ and $\mathcal{T}$. It is often necessary to define the region of interest $\mathcal{R}$ in the source image. The pixels within the region of interest are called the pixels of interest, and their set is written $\mathcal{P}$. It is typical that $\mathcal{R}$ is the convex hull of $\mathcal{P}$. Images are seen as $\mathbb{R}^2 \mapsto \mathbb{R}^c$ functions with $c = 1$ or $c = 3$. Bilinear interpolation is used for non-integer point coordinates. Warps usually have an internal and an external smoothing parameters respectively written λ and μ.

The following acronyms are used:

LLS	Linear Least Squares	NLS	Nonlinear Least Squares
PCA	Principal Component Analysis	RANSAC	Random Sample Consensus
SfM	Structure-from-Motion	SVD	Singular Value Decomposition

There is a strong relationship with curve and surface modeling and fitting tools, as §5.1 emphasizes. A tool that I have not directly made use of is variational calculus. For instance, the cubic spline and the Thin-Plate Spline, which are the solution to variational problems, are used in their parametric forms. I do not use the level set framework (Osher and Paragios, 2003). The parametric and variational frameworks are however strongly linked, and many of the proposed image registration tools could be formulated in the variational framework as well.

CHAPTER

5

DEFORMABLE MODELS AND COMPUTATIONAL METHODS

This chapter brings the context in deformable image models and gives a view to our contributions organized by models and methods. Each section contains a brief review of previous work and the contributions we have made. The first two sections are about modeling, and the last three ones are on computational methods.

We tackle the modeling of image deformations. This includes warps such as Free-Form Deformations and Radial Basis Functions. These warps have an internal built-in smoother. We give our Feature-Driven parameterization for Radial Basis Functions.

We then turn to what we call statistical drivers. These models allow us to incorporate prior knowledge so as to restrict the possible deformations of a warp, and make its estimation better conditioned. We show how the Low-Rank Shape Model allows us to unify all the multilinear statistical drivers, and give our contributions on how it can be efficiently estimated by using a coarse-to-fine ordering on the shape bases.

The first computational methods we examine are the Prediction Sum of Squares statistic and Cross-Validation. They are very useful in the context of estimating empirical deformable models for which there is no statistical model of the residuals.

We give our approach to matrix factorization with missing and erroneous data. This is a difficult Non-linear Least Squares problem. We show that it can be solved using two rounds of optionally robustified convex optimization. This is used for rigid and deformable Structure-from-Motion.

Finally, we investigate the compositional image registration method. This theoretically requires the warp to have a group structure. Our contributions are a means to extend the method to non-groupwise warps, an algorithm that jointly computes the photometric registration and a learning based method for local registration.

5.1 Introduction

The goal of this chapter is to give an overview of the tools we have used and contributed towards and which have had an impact on most of our work. As in most engineering science problems, there are two major steps in solving computer vision problems: a modeling step and an estimation step. The modeling step is to formulate a mathematical model describing the problem and the associated constraints, along with a cost function whose minimum matches the sought after solution. The estimation step is to compute the model parameters from observations by minimizing the cost function. This chapter has two sections on modeling image deformations and three sections on computational estimation methods.

Modeling image deformations. Our study focuses on deformable models. The cameras are modeled by either standard affine or perspective projection. Two aspects are reviewed, the *warps* and what we define as the *drivers*. These two key notions are defined as follows:

- ▷ **Warps** (short for *deformable image warps*) – Example: Thin-Plate Spline (TPS) warps (Bookstein, 1989). We define a warp to be a function that maps an image point to the corresponding point in another image.[1] It is typically driven by a set of deformation centres as for the example of the TPS warp, tuned so that a certain image matching criterion is minimized. A warp is usually dense in the sense that, for any source point, it gives the corresponding target point. It can thus be thought of as an interpolant between the control points. A warp is often based on 2D entities. It often models generic prior knowledge about the problem such as smoothness and rigidity.

- ▷ **Drivers** (short for *statistical deformation driving models*) – Example: 2D Statistical Shape Models (SSM) (Cootes et al., 1991). A driver is what may be used to constrain control points. Another example is the face 3D Morphable Model (3DMM) in (Blanz and Vetter, 1999) which, from a restricted set of parameters, gives the position of some face vertices in 3D which can then be projected with a virtual camera. A control model is in essence sparse, in the sense that, without an interpolation step, it does not take any source point as input. A driver is often based on 3D entities. It models specific knowledge about the problem (*e.g.* as the above mentioned face model does) or very generic knowledge through the Low-Rank Shape Model (LRSM), as is explained later.

Both the warps and drivers can, loosely speaking, be based on 2D or 3D entities and are then said to be 2D or 3D. We show that some very simple warps can be given an interpretation in terms of 3D entities. We rigorously define the lifted affine warps in §5.2.2 as warps that can be written as an unknown projection[2] of some known nonlinearly lifted source point coordinates.

Generally speaking, estimating a 2D warp only solves the matching problem while estimating a 3D driver solves the 3D reconstruction problem. The particular combination of warp and/or driver to use depends on the actual problem setup and images, and on the desired type of outputs. We report some cases of interest:

- ▷ **A deforming surface.** A 2D warp usually manages to register the images, as shown in §6.2.3. One must be careful of the discontinuities induced by phenomena such as self-occlusions, as reported in §6.2.5. If a 3D reconstruction is sought, a 3D driver can be used, either on its own or along with a 2D warp. An example of this is given in §7.1.5. Another option is to use a full 3D warp.

- ▷ **An object of known object-class.** A 3D driver pre-learned from training data typically allows one to register the images and to find a 3D reconstruction, as shown in §7.1.5. The appearance can also be learned. An example of this is the Active Appearance Model (AAM) used in §9.1.

- ▷ **An unstructured deforming environment.** There are not many options in the literature for this case. It is very general since the environment might contain multiple moving and deforming, unsmooth objects. The LRSM gives promising results, as we report in §§7.1.1 to 7.1.4.

[1] These might be two different images of the same scene or two images of similar objects.
[2] This projection is different from the camera projection, though they in general are related.

Some computational methods. We focus on three of the computational methods we have contributed to, since we believe they bring effective solutions to the problems at hand and may be re-used for other classes of problems. We have chosen not to give details about some important computational methods which have recently been successfully applied to the image registration problem. For instance, robust estimators are very important since in feature-based methods there almost always are mismatches, and in pixel-based methods, they are used to handle commonly unmodeled phenomena such as occlusions. Effective methods are the M-estimators (Pilet et al., 2008), RANSAC (Fischler and Bolles, 1981) and its extensions such as PROSAC (Chum and Matas, 2005). A successful approach is to jointly match the points and compute the warp parameters. This is what the TPS-RPM algorithm (Chui and Rangarajan, 2003) and the EM-ICP algorithm (Granger and Pennec, 2002) do. We emphasize the three following computational methods:

- ▷ **Cross-Validation.** For many warp estimation problems, there are weights that need to be tuned. For example, to trade-off the data term and the smoother. It is common to fix these weights by trial and error through visual inspection of the results. We have used criteria such as the Prediction Sum of Squares (PRESS) statistic and Cross-Validation as means to measure the predictivity of the warp or of the model to be estimated. This allows us to estimate the weights and warp parameters automatically and at once.

- ▷ **Matrix factorization with closure and basis constraints.** Matrix factorization appears as a key step in many different problems in computer vision. It is difficult in the case of missing and/or erroneous matrix entries. We report the general solution we have been proposing for rigid and deformable SfM, inspired by (Triggs, 1997b).

- ▷ **Inverse compositional image registration.** This efficient way of registering images, originally proposed in (Baker and Matthews, 2004), is based on the fact that the warp has a group structure on its parameter vector, which is not the case for most deformable image warps. Our contributions are two-fold. First, we have proposed means to approximate the warp inversion and composition for deformable image warps through the Feature-Driven parameterization. Second, we have shown how inverse composition can be used for photometric parameters.

Relationship to curve and surface techniques. Warps are strongly linked to curves and surfaces, as some of those are derived from common curve and surface models such as the cubic spline and its two way tensor product.[3] This makes it difficult to give an extensive overview of the existing possibilities to construct image warps since there is a very large body of literature on curve and surface modeling and fitting (de Boor, 2001; Dierckx, 1993). We thus concentrate on the most widely used approaches in image registration.

Similarly to curves and surfaces, the general idea for estimating warps is to balance the warp complexity while fitting the data. There are basically two main approaches to change the warp complexity: *(i)* change the number of parameters (for instance by adding or removing control points) and *(ii)* penalize the undesired warp behaviour. Both options are studied in the curve fitting literature, see *e.g.* (Dierckx, 1993) for an overview, whereas the second option dominates the image registration literature. We use this approach with Cross-Validation to automatically fit a surface to 2.5D data points in §9.2.2. There however exist successful image registration approaches which automatically tune the number of model parameters such as the Free-Form Deformation (FFD) warp based registration in (Rueckert et al., 1999) or our pixel-based Radial Basis Function (RBF) warp estimation method in §6.2.3.

2.5D and 3D warps and surfaces. There is a strong relationship between the so-called 3D warps and 2.5D surfaces. The reason is that a 3D warp is usually based on combining a 2.5D surface, a camera and possibly the scene flow (Vedula et al., 2005). There are at least two possible constructions for the warps. The first one uses a 2.5D surface, parameterized by an $\mathbb{R}^2 \mapsto \mathbb{R}$ elevation function. This function maps points from the source

[3] The modeling is very similar but the fitting may have two main differences: in image registration, one often has to match the images (it is common in both the pixel- and feature-based approaches that the images must be jointly matched and registered), while in curve and surface fitting, one faces the parameterization problem (that is, the problem of *e.g.* finding a two-dimensional representation for the surface points).

image to their depth. Combining this function with the scene flow and a camera gives the target points. The second possible construction uses a full 3D surface, parameterized by an $\mathbb{R}^2 \mapsto \mathbb{R}^3$ function. This function maps points from a source image to their 3D position, in the coordinate frame of the target camera. The target point is then simply given by dividing the first two coordinates by the third one, representing the depth with respect to the target camera.

One of the Generalized TPS warps we have proposed in §6.2.1 makes use of the second construction to define a 3D deformable warp which incorporates perspective projection effects.

Organization of this chapter. The modeling of image warps is reported in §5.2, which mainly concentrates on 2D warps. Drivers are examined in §5.3, with an emphasize on 3D drivers and the LRSM. Techniques for computing the PRESS and Cross-Validation are given in §5.4. The matrix factorization problem is tackled in §5.5, and inverse compositional image registration in §5.6. A table is given at the beginning of each section. It indicates those of our papers which are included in the companion document. It also quickly summarises the context of the work and the other people who also contributed.

5.2 Deformable 2D Image Warps

The results in this section are mostly related to the following papers:

[I34]	(§6.2.1)	*Generalized Thin-Plate Spline Warps*
[J11]	(§6.2.2)	*Maximizing the Predictivity of Smooth Deformable Image Warps*
[I18]	(§6.2.3)	*Direct Estimation of Non-Rigid Registrations*
[I40]	(§6.2.4)	*Feature-Driven Direct Non-Rigid Image Registration*
[I46]	(§6.2.5)	*Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning*
[I17]	(§7.1.1)	*Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences*
[I29]	(§7.1.5)	*Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model*

I started to work on deformable image warps and registration during my Post Doctoral fellowship with Andrew Zisserman at the University of Oxford. We published papers about the estimation of Thin-Plate Spline warps [I18,I17]. Since then, I have mainly been working on this topic with my two PhD students Vincent Gay-Bellile and Mathieu Perriollat. We published papers proposing the Generalized Thin-Plate Spline warps [I34], giving various methods for the pixel-based estimation of Thin-Plate Spline and Free-Form Deformation warps [I40,I46,I29]. Finally, I have recently proposed a Cross-Validation based method for estimating 2D deformable warps in [J11].

A deformable image warp is an $\mathbb{R}^2 \mapsto \mathbb{R}^2$ function that maps a point in the source image to the corresponding point in the target image. A constructive and a variational approach can be taken to derive image warps. The two approaches are obviously strongly linked. A desirable property for a warp is smoothness. It is then natural to define a warp as the solution to a variational problem with a data term and a smoother.

Organization of this section. First, we give general points about parametric image warps and briefly review the optic flow field representation. Second, we derive the 1D cubic spline and show how both the FFD and TPS warps can be constructed from it, the latter leading to the general RBF warps. Finally, we mention some other warps from the literature.

5.2.1 General Points

5.2.1.1 Parametric Image Warps and General Estimation Scheme

Parametric warps. A parametric image warp maps a point $\mathbf{q}$ from a source image to the corresponding point $\mathbf{q}'$ in the target image. It is written as a function $\mathcal{W} : \mathbb{R}^2 \times \mathbb{R}^p \mapsto \mathbb{R}^2$ of the point coordinates $\mathbf{q} \in \mathbb{R}^2$ and a

parameter vector $\mathbf{u} \in \mathbb{R}^p$ as follows:[4]

$$\mathbf{q}' \;=\; \mathcal{W}(\mathbf{q};\mathbf{u}). \tag{5.1}$$

The parameter vector $\mathbf{u}$ may encapsulate many different kinds of parameters, depending on the nature of the warp, which typically are image control points for 2D warps and 3D control points, surface and camera parameters for 3D warps.

The smoothing-based estimation framework. Generally speaking, the smoothing-based framework is based on minimizing a compound cost-function $\mathcal{E}_c$ that has a *data term* $\mathcal{E}_d$ and an *external smoother* $\mathcal{E}_s$. These two terms can respectively be seen as the likelihood and the prior in the Bayesian context. The problem is thus rewritten:

$$\min_{\mathbf{u}} \mathcal{E}_c^2(\mathbf{u};\mu) \qquad \text{with} \qquad \mathcal{E}_c^2(\mathbf{u};\mu) \;\stackrel{\text{def}}{=}\; \mathcal{E}_d^2(\mathbf{u}) + \mu\mathcal{E}_s^2(\mathbf{u}).$$

The external smoother is weighted by the positive scalar $\mu \in \mathbb{R}^+$. How this scalar can be fixed in practice is examined in §5.4, which describes one of our contributions for the estimation of a warp by minimizing its generalization error. The term $\mathcal{E}_s$ is called the external smoother to make it clear that it differs from the built-in warp smoother, which is called the *internal smoother*. Note that this is slightly different from the meaning used in the active contour or snakes papers (Kass et al., 1988). For instance, an interpolating TPS minimizes the so-called bending energy, modeled through its internal smoother, as we review in §5.2.2.4. It is very common to use as a smoother a penalty on some of the $\gamma \in \mathbb{N}$-th partial derivatives of the warp:

$$\mathcal{E}_{s,\gamma}^2(\mathbf{u}) \;\stackrel{\text{def}}{=}\; \sum_{\mathbf{q}\in\mathcal{R}} \left\| \mathcal{A} \odot \frac{\partial^\gamma \mathcal{W}}{\partial \mathbf{q}^\gamma}(\mathbf{q};\mathbf{u}) \right\|_{\mathcal{F}}^2.$$

In this equation, $\frac{\partial^\gamma \mathcal{W}}{\partial \mathbf{q}^\gamma}(\mathbf{q};\mathbf{u})$ is a rank-γ tensor,[5] and so is $\mathcal{A}$, which gives different weights to the γ-th partial derivatives. Common choices are the first and the second partial derivatives. As an example, the landmark optical flow method in (Horn and Schunck, 1981) uses the first partial derivatives. Other smoothers are possible, such as those dedicated to fluid flow in (Corpetti et al., 2002) or the discontinuity preserving ones (Papenberg et al., 2006).

The second partial derivatives, *i.e.* $\gamma = 2$, with $\mathcal{A} = 1$, *i.e.* an 'all-one' matrix, lead to penalizing the norm of the Hessian matrix $\mathsf{H}(\mathbf{q};\mathbf{u})$ of the warp at points $\mathbf{q} \in \mathcal{R}$ and for parameters $\mathbf{u}$:

$$\mathcal{E}_{s,2}^2(\mathbf{u}) \;\stackrel{\text{def}}{=}\; \sum_{\mathbf{q}\in\mathcal{R}} \|\mathsf{H}(\mathbf{q};\mathbf{u})\|_{\mathcal{F}}^2 \qquad \text{with} \qquad \mathsf{H}(\mathbf{q};\mathbf{u}) \;\stackrel{\text{def}}{=}\; \frac{\partial^2 \mathcal{W}}{\partial \mathbf{q}^2}(\mathbf{q};\mathbf{u}). \tag{5.2}$$

The penalty obviously has to be applied to sufficiently many points to give at least as many independent constraints as there are unknowns. It is usually applied to the pixels in the region of interest $\mathcal{R} \supset \mathcal{P}$.

The data term. The most widely used data terms are roughly speaking *pixel-based* and *feature-based*. Pixel-based[6] means that the value (grey-level or color) of the pixels are directly compared. Feature-based means that an abstraction of the images in terms of some features of interest, typically points or contour curves, is used.

A pixel-based data term is typically written:

$$\mathcal{E}_{dp}^2(\mathbf{u}) \;=\; \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}))\|^2. \tag{5.3}$$

This data term is expressed in pixel value units. It is obviously sensitive to lighting change and pixel color rescaling between the two images. This is dealt with either by making it invariant to such changes or by using a photometric registration model, as discussed below. A simple solution is to replace the pixel values by their

[4] Some warps such as the homography take as input points in $\mathbb{P}^2$.

[5] The rank of a tensor is the number of indices required to describe it. For instance, a scalar is a rank-0 tensor and a matrix is a rank-2 tensor.

[6] This is also called direct and area-based.

spatial derivatives. Assuming that the neighborhood of a pixel is preserved, this achieves invariance to affine color rescaling, see for instance" (Bruhn et al., 2005). Other choices such as Mutual Information are possible. A survey is given in *e.g.* (Pluim et al., 2003). Finally, the function can be robustified using an M-estimator to account for pixels that disappear or which color change is not well modeled by the photometric model. We use such a robustified data term in §6.2.5.

Let $\mathbf{q}_j \in \mathbb{R}^2 \leftrightarrow \mathbf{q}'_j \in \mathbb{R}^2$ be a set of $j = 1, \ldots, m$ point correspondences. A feature-based data term using these point correspondences is the so-called *transfer error*. It is written:

$$\mathcal{E}^2_{df}(\mathbf{u}) \stackrel{\text{def}}{=} \sum_{j=1}^{m} d^2(\mathbf{q}'_j, \mathcal{W}(\mathbf{q}_j; \mathbf{u})).$$

This data term is expressed in pixels. Point matching is usually done first and followed by the estimation of the warp from the point correspondences. A fast and effective approach using a template is described in (Lepetit and Fua, 2006).

We briefly discuss the advantages and drawbacks of the two kinds of data terms. It is generally agreed that feature-based data terms allow computing larger displacements than pixel-based data terms (the so-called *wide-baseline matching* problem). The accuracy reached by the two approaches can be quite different, depending whether the model to be estimated is a global or a local one. For instance, two images of a plane are related by a homography with 8 parameters, which is a global model since the parameters explain the registration for the whole region of interest. In this case, pixel-based and feature-based method will typically reach similar levels of accuracy. For the case of local models, pixel-based methods in general give better accuracy since they use all the possible information available from the images. It seems that threading a robust, wide-baseline feature-based approach and an accurate, pixel-based approach is a sensible approach.

There exist other data terms that are neither feature-based nor pixel-based. Two examples of image-based cues are shading and profiles (or 'occluding contours'), used for instance by (Terzopoulos et al., 1988) for non-rigid 3D reconstruction. They can respectively be related to pixel-based and feature-based data terms, in that shading cues depends on the pixel color, while profiles induce cost functions based on the distance between geometric entities.

Modeling the photometry. Modeling the photometric image variations is required for pixel-based data terms, but also for tasks such as realistic image augmentation. For instance (Luong et al., 2002) shows that the albedoes of a known surface and multiple illuminants can be reconstructed from multiple registered images. This can also be simply done at the image level by estimating a gain and a bias through a 1D affine light change model. We show how an efficient inverse compositional scheme can be used for estimating this model in §6.1.1, and how this generalizes to color images in §6.1.2. A more involved option for image registration is to project the image to an illumination invariant space that gets rid of the light changes and of moving cast shadows, as proposed for a single image in (Finlayson et al., 2006). We adapt this approach to image registration in §6.1.3 and for fitting AAMs in §9.1.2.

5.2.1.2 The Optic Flow Field

The optic flow field is the pixel-wise offsets that must be added to each of the pixel coordinates lying in the region of interest in the source image to get the coordinates of the corresponding pixel in the target image.[7] The parameter vector $\mathbf{u}$ thus contains the offset $\boldsymbol{\delta}_\mathbf{q}$ for each pixel $\mathbf{q} \in \mathcal{R}$, and thus has length $2|\mathcal{R}|$. The optic flow field can be seen in two ways. On the one hand, it is a discrete representation of some image warp at the pixel-level. On the other hand, it can be used in conjunction with some interpolation scheme so that a real point wise image warp is inferred from it. In this case, it is seen as a particular case of an FFD using as a regular grid of control points each pixel in the source image. Interpolation schemes such as pairs of tensor products can be used. More details are given in 5.2.2.3. We note that an optic flow field can also be seen as the natural discretization of the unknown warp in the variational framework.

[7] The flow field representation extends to multiple images, as shown for instance in *e.g.* (Olsen and Nielsen, 2006).

5.2.2 Some Parametric Image Warps

We show how the two most popular parametric warps in the literature, namely FFD and TPS/RBF, can be derived with the 1D cubic spline as a building block. The idea is to build two $\mathbb{R}^2 \mapsto \mathbb{R}$ parametric functions, ϕ_x and ϕ_y, sharing some properties, and then stack them together to form an $\mathbb{R}^2 \mapsto \mathbb{R}^2$ warp ϕ:

$$\phi(\mathbf{q}) \;=\; \begin{pmatrix} \phi_x(\mathbf{q}) \\ \phi_y(\mathbf{q}) \end{pmatrix}.$$

Warps can be roughly classified into 2D and 3D warps. The former are based on image entities only, while the latter uses 3D entities such as cameras.

5.2.2.1 The Lifted Affine Form

We define the *lifted affine warps* as those warps that can be written as the projection of some nonlinearly lifted point coordinates by some unknown projection matrix L:

$$\mathcal{W}_{\mathrm{LA}}(\mathbf{q}; \mathsf{L}) \;\stackrel{\text{def}}{=}\; \mathsf{L}^{\mathsf{T}} \nu(\mathbf{q}). \tag{5.4}$$

The lifting function $\nu : \mathbb{R}^2 \mapsto \mathbb{R}^l$ outputs an l-vector representing the lifted coordinates of a point. They are linearly projected to give the predicted point in $\mathbb{R}^2$ in the target image with the $(l \times 2)$ projection matrix L. As is shown below, this general model encompasses the FFD and RBF warps under common practical assumptions. This is derived through the Feature-Driven parameterizations we have proposed. We show in §6.2.1 that this formulation can be extended to *lifted perspective warps*, allowing us to model perspective projection effects. Thereafter, we assume that L contains the target control points or target deformation centres.

5.2.2.2 The Cubic Spline: A Building Block for Smooth Warps

A spline is a smooth piecewise polynomial function. This name was given in (Schoenberg, 1946). It refers to the flexible device used by draftmen before computers were used for creating engineering drawings. We briefly describe the cubic spline since it allows one to derive the two most popular parametric warps, namely FFD and TPS warps, as two possible extensions of the 1D cubic spline to 2D. We use the B-spline paradigm. Splines are represented as linear combinations of the piecewise cubic polynomial basis functions, the so-called B-splines. This is called "splines in B-form", as coined in (de Boor, 2001). One advantage of the B-splines is that they have local support and minimum degree.

A cubic spline in B-form is a 1D to 1D function parameterized by control points. A detailed derivation following a constructive approach is given in (Dierckx, 1993). Probably more interesting to us is the variational derivation of splines. Consider the variational problem:

$$\min_{\psi} \int_{\mathbb{R}} \left(\frac{\mathrm{d}^2 \psi}{\mathrm{d}x^2} \right)^2 \mathrm{d}x \qquad \text{s.t.} \qquad \psi(x_k) = z_k, \;\; k = 1, \ldots, \eta.$$

The cost functional is a smoother which penalizes the curvature of the spline while the constraints make it go through the data points (x_k, z_k). It has been shown that the solution is the cubic spline, with a sequence of knots that coincide with the data. The cubic degree comes from the smoother which uses second derivatives. This variational derivation clearly shows the analogy with the physical spline device: the constraints represent the anchor points to which the flexible strip is attached, while the smoother is an approximation to its *bending energy*.

The cubic spline in B-form is:

$$\psi(x) \;=\; \sum_{k=1}^{\eta} z_k B_k(x), \tag{5.5}$$

with B_k the k-th cubic B-spline. This is sometimes called the interpolating spline, as opposed to the smoothing spline, which minimizes the following functional, balancing goodness of fit and smoothness:

$$\min_{\psi} \sum_{k=1}^{\eta} (z_k - \psi(x_k))^2 + \lambda \int_{\mathbb{R}} \left(\frac{\mathrm{d}^2 \psi}{\mathrm{d}x^2} \right)^2 \mathrm{d}x. \tag{5.6}$$

The solution again is a cubic spline with knots at the data points and with control points estimated through LLS.

In the following, we use equally-spaced knot sequences, leading to uniform splines, for which the B-spline functions are translated versions of each other. The l knots are placed such that the inter-knot distance is unity. The blending function can be pre-calculated and gives:

$$\psi(x) \;=\; \sum_{a=0}^{3} B_a(x - \lfloor x \rfloor) z_{\lfloor x \rfloor + a},$$

with:

$$\begin{array}{llll} B_0(x) & \stackrel{\text{def}}{=} & \frac{1}{6}(-x^3 + 3x^2 - 3x + 1) & \qquad B_1(x) \stackrel{\text{def}}{=} \frac{1}{6}(3x^3 - 6x^2 + 4) \\ B_2(x) & \stackrel{\text{def}}{=} & \frac{1}{6}(-3x^3 + 3x^2 + 3x + 1) & \qquad B_3(x) \stackrel{\text{def}}{=} \frac{1}{6}x^3. \end{array}$$

The $\mathbb{R} \mapsto \mathbb{R}$ cubic spline in B-form is driven by the scalars z_k defined at each knot, that are linearly combined. It is straightforward to extend these splines to $\mathbb{R} \mapsto \mathbb{R}^d$ with $d \geq 1$, by replacing the scalars z_k by *target control points* in $\mathbb{R}^d$. One of the advantages of these functions is their localized support: the value at a point only depends on 4 neighboring control points.

5.2.2.3 Free-Form Deformation Warps

$\mathbb{R}^2 \mapsto \mathbb{R}$ Free-Form Deformations. The FFD warps are based on the tensor product between two $\mathbb{R} \mapsto \mathbb{R}$ splines. The *source control points* thus lie on a regular grid. It was proposed in (Sederberg and Parry, 1986) for computer graphics applications and has been extensively used since then. Early work with FFD warps for image registration is (Rueckert et al., 1999; Szeliski and Coughlan, 1997). Dirichlet FFDs, proposed in (Moccozet and Magnenat-Thalman, 1997) and used in (Ilíc and Fua, 2002) for deformable model fitting, do not require that the source control points lie on a regular grid. More recently, FFD warps have been used for shape registration in (Huang et al., 2006).

Consider two one-dimensional sets of evenly spaced knots with unity inter-knot distance. Assume one set spans the horizontal, x axis of the image, and the other one spans the vertical, y axis. These two sets of knots define a regular grid whose vertices are the source control points. A scalar target value $z_{u,v}$ is associated to each source control point $(u \;\; v)^{\mathsf{T}} \in \mathbb{N}^2$.

For a point $\mathbf{q} \in \mathbb{R}^2$ with $\mathbf{q}^{\mathsf{T}} = (x \;\; y)$, the tensor product is written:

$$\sum_{a=0}^{3} \sum_{b=0}^{3} B_a(x - \lfloor x \rfloor) B_b(y - \lfloor y \rfloor) z_{\lfloor x \rfloor + a, \lfloor y \rfloor + b}.$$

The $\mathbb{R}^2 \mapsto \mathbb{R}^2$ warp is obtained by stacking two such tensor products sharing their source control points, or equivalently, by replacing the scalars $z_{u,v}$ by so-called *target control points* $\mathbf{c}_{u,v}$, giving the FFD warp as:

$$\mathcal{W}_{\mathrm{FFD}}(\mathbf{q}; \mathsf{L}) \;\stackrel{\text{def}}{=}\; \sum_{a=0}^{3} \sum_{b=0}^{3} B_a(x - \lfloor x \rfloor) B_b(y - \lfloor y \rfloor) \mathbf{c}_{\lfloor x \rfloor + a, \lfloor y \rfloor + b}. \tag{5.7}$$

Incremental Free-Form Deformation warps. It often is the case that the source control points are known, and the target control points are unknown. The former can for instance be chosen so that they cover the region of interest in the source image. It is sometimes convenient to use as unknowns the displacements $\boldsymbol{\delta}_{u,v}$ between the source and the target control points:

$$\boldsymbol{\delta}_{u,v} \;\stackrel{\text{def}}{=}\; \mathbf{c}_{u,v} - \begin{pmatrix} u \\ v \end{pmatrix}.$$

The Incremental FFD warp is a rewriting of the FFD warp (5.7) in terms of the displacements:

$$\mathcal{W}_{\mathrm{IFFD}}(\mathbf{q};\mathsf{L}) \stackrel{\text{def}}{=} \mathbf{q} + \sum_{a=0}^{3}\sum_{b=0}^{3} B_a(x-\lfloor x\rfloor)B_b(y-\lfloor y\rfloor)\boldsymbol{\delta}_{\lfloor x\rfloor+a,\lfloor y\rfloor+b}.$$

This property stems from the fact that the source control points give the identity transformation when used as target control points, since a B-spline through colinear control points in a straight line.

Hierarchical Free-Form Deformation warp. So as to ease convergence to the right solution, it is typical to embed FFD warps in a coarse-to-fine framework, where the number of control points is progressively increased as the fitting proceeds, as *e.g.* in (Huang et al., 2006; Rueckert et al., 1999). A spline subdivision algorithm such as the one in (Weimer and Warren, 1998) is used to refine the lattice of control points to the finer level.

The lifted affine form. We show that the FFD is a lifted affine warp in the sense of equation (5.4) by rewriting it as:

$$\mathcal{W}_{\mathrm{FFD}}(\mathbf{q};\mathsf{L}) = \mathsf{L}^{\mathsf{T}}\nu_{\mathrm{FFD}}(\mathbf{q}).$$

The l vector $\nu_{\mathrm{FFD}}(\mathbf{q})$, *i.e.* the nonlinearly lifted coordinates, is defined as the 16 tensor product cubic B-spline coefficients placed appropriately. We point out that this is not a Feature-Driven parameterization *stricto sensu* since an FFD does not interpolate its control points.

5.2.2.4 Thin-Plate Spline and Radial Basis Function Warps

The TPS equation as the solution to a variational problem was shown in (Duchon, 1976) while the proof of uniqueness was later established in (Wahba, 1990). It was first used to construct image warps in (Bookstein, 1989).

The $\mathbb{R}^2 \mapsto \mathbb{R}^1$ Thin-Plate Spline. Alternatively to using the tensor product to extend the splines to $\mathbb{R}^2 \mapsto \mathbb{R}$ functions, one can consider extending to 2D and solving the variational problem (5.6) from which the splines are derived. This variational problem can be extended as:

$$\min_{\zeta} \sum_{k=1}^{l}(z_k - \zeta(\mathbf{b}_k))^2 + \lambda \int_{\mathbb{R}^2} \left\| \frac{\partial^2 \zeta}{\partial \mathbf{q}^2}(\mathbf{q}) \right\|_{\mathcal{F}}^2 \mathrm{d}\mathbf{q}, \tag{5.8}$$

where l is the number of deformation centres. The points $\mathbf{b}_k$ are called *source deformation centres* or just source centres. This variational problem has an analytical solution: the TPS. This way of deriving the TPS explains why it is sometimes called the "natural extension of the 1D cubic spline to 2D". The TPS is written:[8]

$$\varphi(\mathbf{q};\boldsymbol{\xi}_{\mathbf{z};\lambda}) \stackrel{\text{def}}{=} \mathbf{a}^{\mathsf{T}}\tilde{\mathbf{q}} + \sum_{k=1}^{l} \varrho(\|\mathbf{q}-\mathbf{b}_k\|_2^2)w_k, \tag{5.9}$$

with $\tilde{\mathbf{q}}$ the homogeneous coordinates of point $\mathbf{q}$. We observe that the TPS has two parts: an affine part with 3 coefficients in $\mathbf{a}$, and a radial basis part with coefficients w_k, gathered into $\boldsymbol{\xi}_{\mathbf{z},\lambda}^{\mathsf{T}} \stackrel{\text{def}}{=} (\mathbf{w}^{\mathsf{T}} \ \mathbf{a}^{\mathsf{T}})$. The function ϱ is the TPS basis function for the square distance, given by:

$$\varrho(d^2) \stackrel{\text{def}}{=} d^2\log(d^2).$$

[8]The solution $\zeta^\star$ to (5.8) is an $\mathbb{R}^2 \mapsto \mathbb{R}^2$ function but we write it as an $\mathbb{R}^2 \times \mathbb{R}^{l+3} \mapsto \mathbb{R}^2$ function φ to emphasize its dependency on the an $l+3$ coefficient vector $\boldsymbol{\xi}_{\mathbf{z},\lambda}$ subject to the side-conditions.

The coefficients are computed by solving a linear system of equations, obtained by writing the interpolation constraints:

$$\underbrace{\begin{pmatrix} \mathsf{K}_\lambda & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & 0 \end{pmatrix}}_{\mathcal{D}} \underbrace{\begin{pmatrix} \mathbf{w} \\ \mathbf{a} \end{pmatrix}}_{\boldsymbol{\xi}_{\mathbf{z},\lambda}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{0} \end{pmatrix} \quad \text{with} \quad K_{r,k} \stackrel{\text{def}}{=} \begin{cases} \lambda & r = k \\ \varrho(\|\mathbf{b}_r - \mathbf{b}_k\|^2) & \text{otherwise,} \end{cases} \tag{5.10}$$

with $\mathsf{B}^\mathsf{T} \stackrel{\text{def}}{=} (\tilde{\mathbf{b}}_1 \ \cdots \ \tilde{\mathbf{b}}_l)$. The last three equations in $\mathsf{B}^\mathsf{T}\mathbf{w} = \mathbf{0}$ are called 'side-conditions'. They ensure that the TPS has square integrable second derivatives. The square integral bending energy is given by:

$$\kappa \stackrel{\text{def}}{=} \int_{\mathbb{R}^2} \left\| \frac{\partial^2 \varphi}{\partial \mathbf{q}^2}(\mathbf{q}; \boldsymbol{\xi}_{\mathbf{z},\lambda}) \right\|_{\mathcal{F}}^2 \mathrm{d}\mathbf{q} = 8\pi \mathbf{w}^\mathsf{T} \mathsf{K}_\lambda \mathbf{w}. \tag{5.11}$$

The $\mathbb{R}^2 \mapsto \mathbb{R}^2$ Thin-Plate Spline warp. Stacking two such TPS sharing their centres gives the $\mathbb{R}^2 \mapsto \mathbb{R}^2$ TPS warp:

$$\mathcal{W}_{\mathrm{TPS}}(\tilde{\mathbf{q}}; \Xi_{\mathsf{L};\lambda}) \stackrel{\text{def}}{=} \mathsf{A}\mathbf{q} + \sum_{k=1}^{l} \varrho(\|\mathbf{q} - \mathbf{b}_k\|_2^2)\mathbf{w}_k, \tag{5.12}$$

where $\Xi_{\mathsf{L};\lambda}$ is a two column matrix gathering the coefficients in vector $\boldsymbol{\xi}_{\mathbf{z},\lambda}$ for the x and y axes. One advantage of TPS warps over FFD warps is that the source centres can be placed anywhere in the image. The disadvantage is clearly that the TPS kernel ϱ has a global support, leading to dense matrices in the computation procedures.

Radial Basis Functions. It can be seen that the TPS (5.9) is an RBF[9] since, omitting its affine counterpart, it depends only on the distance between the point at which it is evaluated and each of the source centres. The TPS is the RBF that minimizes the bending energy. There exist many different RBFs which are the solution to variational problems of the form (5.8) with different smoothers. These problems are solved within the framework of Reproducing Kernel Hilbert Spaces (RKHS) and has been extensively studied, as for instance in the book (Wahba, 1990).

The Feature-Driven parameterization. The FFDs of §5.2.2.3 are naturally driven by their target control points while, as derived above, the TPS is parameterized by $l + 3$ coefficients in $\boldsymbol{\xi}_{\mathbf{z},\lambda}$, see equation (5.9). Directly optimizing over these $l + 3$ coefficients has several disadvantages: one has to ensure that the side-conditions are satisfied, and these coefficients do not have explicit units. Regarding the affine counterpart of the TPS warp (5.12), we recall that the first two columns of matrix A can be interpreted as direction vectors, so are unitless, while the third one represents an offset and is thus in pixels. This definitely is not a well-balanced parameterization.

We propose the Feature-Driven parameterization of TPS warps in §6.2.1 and summarise it below. This enables driving these by the coordinates of *target deformation centres*, or just target centres, similarly to the FFD warps being driven by target control points. The Feature-Driven parameterization forms the basis for the Generalized TPS warps we propose in §6.2.1. This is related to the parameterization proposed in (Lim and Yang, 2005).

First, we define the $(l + 3)$-vector $\boldsymbol{\ell}_\mathbf{q}$ as:

$$\boldsymbol{\ell}_\mathbf{q}^\mathsf{T} \stackrel{\text{def}}{=} (\varrho(\|\mathbf{q} - \mathbf{b}_1\|_2^2) \ \cdots \ \varrho(\|\mathbf{q} - \mathbf{b}_l\|_2^2) \ \tilde{\mathbf{q}}^\mathsf{T}),$$

allowing us to rewrite the TPS (5.9) as a dot product:

$$\varphi(\mathbf{q}) = \boldsymbol{\ell}_\mathbf{q}^\mathsf{T} \eta_{\mathbf{z},\lambda}. \tag{5.13}$$

[9] A Radial Basis Function is defined as a sum of radially symmetric functions, each of them depending only on the distance between the source point and a centre.

Second, we express $\eta_{\mathbf{z},\lambda}$ as a linear 'back-projection' of the target value vector $\mathbf{z}$. This is modeled by the matrix X_λ, nonlinearly depending[10] on λ, given by the l leading columns of $\mathcal{D}^{-1}$ from equation (5.10). Some algebraic manipulations show that:

$$\boldsymbol{\xi}_{\mathbf{z},\lambda} \;=\; \mathsf{X}_\lambda \mathbf{z} \qquad \text{with} \qquad \mathsf{X}_\lambda \;\stackrel{\text{def}}{=}\; \begin{pmatrix} \mathsf{K}_\lambda^{-1}\left(\mathrm{I} - \mathsf{B}\left(\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1}\mathsf{B}\right)^{-1}\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1}\right) \\ \left(\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1}\mathsf{B}\right)^{-1}\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1} \end{pmatrix}. \tag{5.14}$$

This parameterization has the advantage of separating λ and $\mathbf{z}$, and to introduce units.[11] It also naturally enforces the side-conditions.

Incorporating the parameterization (5.14) into the TPS (5.13) we obtain our Feature-Driven parameterization $\tau(\mathbf{q};\mathbf{z},\lambda) = \varphi(\mathbf{q};\boldsymbol{\xi}_{\mathbf{z},\lambda})$:

$$\tau(\mathbf{q};\mathbf{z},\lambda) \;\stackrel{\text{def}}{=}\; \boldsymbol{\ell}_\mathbf{q}^\mathsf{T}\mathsf{X}_\lambda\mathbf{z}. \tag{5.15}$$

With this parameterization, the bending energy (5.11) is rewritten as:

$$\kappa \;=\; 8\pi\mathbf{z}^\mathsf{T}\bar{\mathsf{X}}_\lambda\mathbf{z} \qquad \text{with} \qquad \bar{\mathsf{X}}_\lambda \;\stackrel{\text{def}}{=}\; \mathsf{K}_\lambda^{-1}\left(\mathrm{I} - \mathsf{B}\left(\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1}\mathsf{B}\right)^{-1}\mathsf{B}^\mathsf{T}\mathsf{K}_\lambda^{-1}\right). \tag{5.16}$$

Matrix $\bar{\mathsf{X}}_\lambda$ is the $(l\times l)$ *bending energy matrix* given by truncating X_λ by the last three rows. The bending energy matrix is symmetric and in the absence of internal regularization, *i.e.* for $\lambda = 0$, has rank $l-3$. The eigenvectors corresponding to the $l-3$ nonzero eigenvalues are the *principal warps*, the corresponding eigenvalues indicating their bending energy (Bookstein, 1989).

The TPS warp is obtained by stacking two $\mathbb{R}^2 \mapsto \mathbb{R}$ TPSs. From equation (5.15), we get:

$$\mathcal{W}_{\mathrm{TPS}}(\mathbf{q};\mathsf{L}) \;=\; \begin{pmatrix}\tau(\mathbf{q};\boldsymbol{\alpha}_x,\lambda)\\ \tau(\mathbf{q};\boldsymbol{\alpha}_y,\lambda)\end{pmatrix} \;=\; \left(\boldsymbol{\ell}_\mathbf{q}^\mathsf{T}\mathsf{X}_\lambda\mathsf{L}\right)^\mathsf{T} \;=\; \mathsf{L}^\mathsf{T}\mathsf{X}_\lambda^\mathsf{T}\boldsymbol{\ell}_\mathbf{q}, \tag{5.17}$$

where $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ are the first and second columns of L respectively.

Finally, we can use the following smoother as an equivalent to the discretized bending energy (5.2):

$$\kappa \;=\; 8\pi\|\mathsf{Z}\mathsf{L}\|_\mathcal{F}^2,$$

with matrix Z chosen such that $\mathsf{Z}^\mathsf{T}\mathsf{Z} = 8\pi\bar{\mathsf{X}}$. Note that in practice, one does not need to compute Z since only $\mathsf{Z}^\mathsf{T}\mathsf{Z}$ is needed, *e.g.* for building the influence matrix needed to cross-validate the warp, as we explain in §5.4. The Feature-Driven parameterization can be similarly derived for all RBFs.

The lifted affine form. It is straightforward from the Feature-Driven parameterization (5.17) to express the TPS warp in the lifted affine form (5.4), *i.e.* $\mathcal{W}_{\mathrm{TPS}}(\mathbf{q};\mathsf{L}) = \mathsf{L}^\mathsf{T}\nu_{\mathrm{TPS}}(\mathbf{q})$, with the following nonlinear lifting function:

$$\nu_{\mathrm{TPS}}(\mathbf{q}) \;=\; \mathsf{X}_\lambda^\mathsf{T}\boldsymbol{\ell}_\mathbf{q}.$$

5.2.2.5 Summary and a Short Comparison

We have reviewed three different kinds of 2D warp parameterization: the flow-field, the FFD and the RBF, which includes the TPS. Which one to choose in practice depends on the problem at hand. The major advantage of the flow-field is its extreme locality: each parameter influences only one or two measurements, leading to highly sparse design matrices. The flow-field must however be interpolated by some smooth kernel in order to map points with real coordinates. The most natural interpolation scheme is actually an FFD with every pixels as control points. The flow-field can be seen as a natural discretization of a variational warp estimation problem.

The FFD too has the advantage of locality. For example, when a cubic spline is used as a basis for the tensor product, a point is influenced only by the 16 surrounding control points. This approach however is limited, at least in its basic form, by the fact that the control points are constrained to lie on a grid.

[10] The internal smoothing parameter λ is chosen small to ensure that matrix X_λ is well-conditioned.

[11] While $\boldsymbol{\xi}_{\mathbf{z},\lambda}$ has no obvious unit, $\mathbf{z}$ in general has (*e.g.* pixels, meters).

The RBF has the advantage that its deformation centres can be located arbitrarily.[12] However, the major problem with RBFs is finding the right kernel (and its parameters). Using a localized kernel in general is difficult since there is not a universally agreed method for choosing the kernel width; one of the most important kernel parameters. On the other hand, the TPS kernel is in general very well behaved, but has an infinite support. It has a stiffness parameter but which in practice has a very limited influence on the resulting warps. We note that localized approximations to the TPS have also been proposed in for example (Donato and Belongie, 2002; Fornefett et al., 2001; Zandifar et al., 2004).

5.2.3 Other Kinds of Warps

Although the FFD and TPS warps are the most common smooth warps, there are numerous other warps used in the literature. Some of those are briefly reviewed below. A simple solution is to construct a piecewise affine warp by fitting a triangular mesh to control points. Each triangle defines an affine transformation. This is typically used in AAM based registration such as in (Cootes et al., 2001), in the registration framework in (Pilet et al., 2008) or in the deformable shape detection algorithm in (Felzenszwalb, 2003).

The generalized elastic nets are described in (Myronenko et al., 2007). The idea is to represent the source image by a constrained Gaussian mixture that is fitted to the target image. Other examples are elastic registration which uses spring terms (Christensen and He, 2001) and fluid registration which uses viscosity (Bro-Nielsen and Gramkow, 1996). Brownian warps are proposed in (Nielsen et al., 2002) along with a smoother constraining the estimated warp to be invertible (Nielsen and Johansen, 2004). Similarly, the groupwise warps in (Cootes et al., 2004) are constructed to be invertible.

There are numerous deformation models from the computer graphics, computer-aided design and statistical data modeling communities that we can use. Basically, most curve modeling and deformation methods, and surface models, can be extended to give image warps, in a similar fashion to how FFD and RBF warps are constructed above. These are usually used for shape interpolation and not for registration. We name just a few: the Moving Least Squares warps (Schaefer et al., 2006), and some of the possible extensions of the 1D cubic spline to 2D such as the splines over triangular meshes (Powell and Sabin, 1977). As-rigid-as-possible warps are presented in (Alexa et al., 2000), while (Weng et al., 2006) demonstrate global and local shape preserving nonlinear transformations.

[12] Except for some kernels for which colinear centres induce degeneracies – this can be dealt with by using a light-weighted internal smoother.

5.3 The Low-Rank Shape Model and Other Statistical Drivers

The results in this section are mostly related to the following papers:

[I17]	(§7.1.1)	*Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences*
[I22]	(§7.1.2)	*A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery*
[J10]	(§7.1.3)	*Implicit Non-Rigid Structure-from-Motion with Priors*
[I50]	(§7.1.4)	*Coarse-to-Fine Low-Rank Structure-from-Motion*
[I29]	(§7.1.5)	*Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model*
[I25]	(§7.2.2)	*Towards 3D Motion Estimation from Deformable Surfaces*
[I41]	(§9.1.1)	*Segmented AAMs Improve Person-Independent Face Fitting*
[I49]	(§9.1.2)	*Light-Invariant Fitting of Active Appearance Models*

I have first used the implicit Low-Rank Shape Model during my Post Doctoral fellowship with Andrew Zisserman at the University of Oxford. We have used it to constrain a Thin-Plate Spline warp between images estimated from curve correspondences [I17]. I have extended it with my PhD students Vincent Gay-Bellile and Mathieu Perriollat in [I29] so as to incorporate synthetically generated training data. I have used this model to compute the pose of a 3D sensor in [I25]. I have then defined, with Søren Olsen from the University of Copenhagen, an implicit Low-Rank Shape Model [I22,J10]. Finally, I have collaborated with Julien Peyras from the University of Milan on the face Active Appearance Model fitting problem [I41]. We have recently published a paper with Daniel Pizarro from the University of Alcala, proposing a light-invariant fitting method [I49]. My most recent contribution is a coarse-to-fine explicit Low-Rank Shape Model [I50].

Image warps as described in the previous section may have a large number of parameters, as for instance the flow-field where the number of parameters is twice the number of pixels. The general purpose of statistical drivers is to reduce this number of parameters by embedding prior knowledge. This should reflect the dependencies between the point displacements, so that they can be parameterized by fewer parameters. There is a great deal of work on physically based models (Metaxas, 1997). Examples are the snakes which deform elastically (Kass et al., 1988), the vibrational models in (Park et al., 1996; Pentland and Sclaroff, 1991) or the nonlinear beam model in (Ilíc and Fua, 2007). A recent approach is Generalized PCA (Vidal et al., 2005). We are interested here in the multilinear drivers described below, mainly for their flexiblity and representational power.

We consider that the object or the class of the object of interest, for example a face, is known. One of the first approaches one may think about is to train a driver using PCA on a set of 2D training shapes representing the extent to which the shape can deform. The classical example for this is the SSM proposed in (Cootes et al., 1991) which is usually fitted to a single image for detection, localization and segmentation purposes. This approach is very effective in some cases but cannot be readily used to get 3D information. It may encounter difficulties with for instance pose and lighting variations. This led to the development of the 3DMM proposed in (Blanz and Vetter, 1999). A 3DMM is constructed by using PCA on 3D training data. At the fitting stage, the 3D model parameters and unknown camera pose are recovered.

The major limitation of the above drivers is that they do not cope with general scenes. Recently, researchers have proposed to make these drivers more flexible. Considering multiple images without knowing what kind of scene is observed, they learn the PCA model directly from the actual data. The world that surrounds us is indeed highly structured, but its structure is highly complicated. The hope for statistical drivers is that they manage to capture this structure. PCA, as one of the mostly used dimensionality reduction algorithms, is a natural way to describe that the data has some structure, even though this structure has to be discovered. It can be seen as a generic prior on the scene structure. This approach, pioneered in (Irani, 1999) has been very successful in 2D. It has been extended to the 3D case in (Bregler et al., 2000) and shows to be very promising. It is called the explicit LRSM for reasons that are made clear below. The Low-Rank assumption is that the rank of the measurement matrix, related to the number of shape bases needed to describe the deformations, is low compared to the amount of data (*i.e.* the number of views and the number of point tracks).

These drivers lie in a class we coined *multilinear drivers*. Indeed, considering that a driver-generated shape

is a set of points,[13] PCA-like models express this shape as a linear combination of *shape bases*. The shape is thus a linear function of the shape bases and of the *configuration weights*. This strictly holds for the 2D drivers and for those 3D drivers using an affine camera model. Slightly abusing the expression, we assume it also includes the 3D drivers using the perspective camera model, which is linear in the homogeneous point coordinates.

We distinguish the following characteristics for statistical drivers, stemming from the previous discussion:

- ▷ **2D versus 3D.** A 3D driver combines a camera model with a 3D deformable shape. It has the advantage that it may allow one to recover the 3D structure and camera pose, but is in general 'more nonlinear' than a 2D driver.

- ▷ **Pre-trained versus un-trained.** A pre-trained driver is dedicated to a specific object or object-class, whereas an untrained driver is trained online, *i.e.* on the actual data. A pre-trained driver is thus more specific and better posed in terms of 3D structure recovery.

We focus on multilinear drivers generating a set of points to represent the shape, and study the 2D and 3D, pre-trained and un-trained cases. This is summarised in the following table:

	2D	3D
Pre-trained	SSM (Statistical Shape Models) linear	3DMM (3D Morphable Models) at least bilinear
Un-trained	Implicit LRSM (Low-Rank Shape Model) bilinear	Explicit LRSM (Low-Rank Shape Model) at least trilinear

From this table, it is clear that the explicit LRSM is the most general driver, and provides a unified formulation from which the other, more specific drivers can be derived.

One of the aims in using the explicit LRSM driver is *Monocular Deformable SfM in a generic manner*. Most existing systems such as (White et al., 2007) are based on multiple cameras and/or specific patterns. The LRSM driver proceeds by projecting linearly combined, un-trained 3D shape bases. This is actually an ill-posed problem, and it is indeed not clear at first sight how a simple multilinear model using very few and generic prior assumptions may be able to reconstruct the camera motion and scene structure. The most common estimation approach is the stratified one: first, estimating the 2D, implicit LRSM driver, which is then used for estimating the 3D, explicit one. It turns out that using more priors such as continuity or even smoothness of the camera path and orientation have recently been shown to be of crucial importance, as we report in §7.1.3 and as (Torresani et al., 2007) demonstrates. Another aspect we believe to be very important is that the estimation method is rooted in a proper definition of what the motion and deformation are in this context. One definition for these notions is given in the Deformotion paper (Yezzi and Soatto, 2003). Loosely speaking, they define the object motion as the best fit with respect to a mean shape and to a motion group, and the object deformation as the residuals. This inspired us for the coarse-to-fine reconstruction method we propose in §7.1.4.

Finally, one must consider the problem of selecting the number of shape bases, also known as the rank selection problem. This is a difficult problem since a multilinear driver is empirical: it is not derived from a physical deformation model, but rather attempts at capturing some structure from the data. This is made even more difficult if the training data may be partly erroneous, since one has to face the ambiguous problem of distinguishing deformations from blunders.

Organization of this section. In the remainder of this section we first describe a selection of pre-trained drivers, and second, the un-trained LRSM driver. We derive the 2D and 3D counterparts for each case. Third, we review some generic priors. Fourth, we describe ways for estimating the 2D and 3D LRSM parameters leading to Low-Rank SfM. Finally, we briefly mention the rank selection problem and possible extensions.

[13] Many drivers such as the AAM (Cootes et al., 2001) and the 3DMM (Blanz and Vetter, 1999) also learn and generate an appearance counterpart.

5.3.1 Pre-Trained Drivers

Training a driver means estimating its shape bases from a set of registered data. Pre-trained drivers are thus specific to an object or an object-class. They can be fitted to a single image and in the 3D case, are usually constrained enough to recover a decent 3D reconstruction, and can even be used for multiple camera calibration (Koterba et al., 2005). These drivers are also used for motion synthesis as for example in (Urtasun et al., 2004). Recall that we are interested in using these as warp drivers; *i.e.* for fitting them to several monocular images at once to guide a warp.[14] They are 'more linear' than the un-trained models.

The training step is in principle similar as the reconstruction of an un-trained driver. Since it is done only once, it usually is achieved with data that specifically ease the training process. In other words, the training and test data are not of the same nature. For instance, the face 3DMM in (Blanz and Vetter, 1999) is trained from registered 3D face scans, even though it is to be fitted to standard face images.

5.3.1.1 The 3D Case: 3D Morphable Models

A 3DMM describes the 3D shape points as a linear combination of l pre-learned 3D shape bases $\mathbf{B}_{k,j} \in \mathbb{R}^3$ with configuration weights $\alpha_k \in \mathbb{R}$. The image points $\mathbf{q}_j \in \mathbb{R}^2$ are obtained by projecting the 3D shape points using a projection operator $\Pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$ modeling the camera:

$$\mathbf{q}_j \;=\; \Pi\left(\sum_{k=1}^{l} \alpha_k \mathbf{B}_{k,j}\right). \tag{5.18}$$

This equation is trilinear if an affine camera model is used for Π. The 3DMM has been popularized by the face model in (Blanz and Vetter, 1999) where the shape is learned along with an appearance counterpart. It allows one to find the 3D shape of a face from a single image. This driver is specific since it is dedicated to faces, but models a broad range of faces.

It has recently been used as a means to reduce the number of parameters in continuous surface modeling (Salzmann et al., 2007b). The approach they took is to learn the model from synthetically generated training data. The driver can be used for many different kinds of surfaces, and is in this sense generic. We used this model as a 3D driver combined to a TPS warp in §7.1.5. The overall warp is a 3D warp.

5.3.1.2 The 2D Case: Statistical Shape Models

SSMs can be seen as 2D drivers since they are based on 2D shape bases $\mathbf{b}_{k,j} \in \mathbb{R}^2$ and synthesize the image points $\mathbf{q}_j \in \mathbb{R}^2$ with the following bilinear model:

$$\mathbf{q}_j \;=\; \sum_{k=1}^{l} \alpha_k \mathbf{b}_{k,j}.$$

The 2D shape bases are learned by PCA on annotated training shapes. SSMs were proposed by Cootes *et al.* (Cootes et al., 1991). They can be used in conjunction with an appearance model to form the AAMs. They are often used to model body parts in medical image analysis and faces in computer vision. The representational power of SSMs is studied in (Xiao et al., 2004), where it is shown that under some hypotheses, they can be as expressive as 3D AAMs. Indeed, using an affine camera model for the projection operator Π in the 3DMM driver (5.18) shows that the 2D shape bases $\mathbf{b}_{k,j}$ can be seen as projected 3D shape bases $\mathbf{B}_{k,j}$. We have contributed to the AAM face model in §9.1.1 by proposing a fitting strategy for segmented AAMs and a statistical error measure, and in §9.1.2 with a light invariant fitting approach dealing with the difficult issues of external and self-shading.

[14] When one uses a template as one of the images to register, the problem then is like the one of fitting an appearance modeling driver such as an AAM with no appearance variation mode.

5.3.2 Un-Trained Drivers

A paradigm has recently emerged in computer vision: the one of training a 3DMM from the actual image data. This is called the LRSM, and has been pioneered in (Bascle and Blake, 1998; Boult and Brown, 1991; Brand, 2001; Bregler et al., 2000; Irani, 1999; Torresani et al., 2001).

Contrarily to the pre-trained drivers, the un-trained ones have to discover the data structure and regularities from a single dataset. They are thus useful for dealing with scenes consisting of single or multiple unidentified objects present in multiple images.

5.3.2.1 The 3D Case: The Explicit Low-Rank Shape Model

The explicit LRSM writes as the 3DMM (5.18). Introducing a frame index i, an image point $\mathbf{q}_{i,j} \in \mathbb{R}^2$ is given by:

$$\mathbf{q}_{i,j} = \Pi_i \left(\sum_{k=1}^{l} \alpha_{i,k} \mathbf{B}_{k,j} \right), \tag{5.19}$$

where $\alpha_{i,k} \in \mathbb{R}$ are view dependent configuration weights. The concept of learning the shape bases from the actual data makes the driver much more flexible and generic but raises the question of whether the model is well-posed, in the sense that it has a unique solution (up to some gauge transformation) that matches reality. Estimating the parameters in (5.19) is a difficult problem for which solutions are reviewed in §5.3.4.

5.3.2.2 The 2D Case: The Implicit Low-Rank Shape Model

The implicit LRSM is the 2D counterpart of the explicit one, and is derived from (5.19). Assuming an affine projection model with (2×3) rotational parts P_i and (2×1) translational parts $\mathbf{t}_i$, (5.19) is rewritten:

$$\mathbf{q}_{i,j} = \mathsf{P}_i \sum_{k=1}^{l} \alpha_{i,k} \mathbf{B}_{k,j} + \mathbf{t}_i.$$

Moving P_i inside the summation and re-arranging gives:

$$\mathbf{q}_{i,j} = \begin{pmatrix} \alpha_{i,1}\mathsf{P}_i & \cdots & \alpha_{i,l}\mathsf{P}_i \end{pmatrix} \begin{pmatrix} \mathbf{B}_{1,j} \\ \vdots \\ \mathbf{B}_{l,j} \end{pmatrix} + \mathbf{t}_i = \mathsf{M}_i \mathbf{S}_j + \mathbf{t}_i, \tag{5.20}$$

where M_i and $\mathbf{S}_j$ have size $(2 \times r)$ and $(r \times 1)$ respectively. $r = 3l$ is the so-called *driver rank* since it corresponds to the rank of the data matrix in the factorization formulation described in §5.3.4. We next introduce an $(r \times r)$ full-rank mixing matrix E which yields the implicit LRSM driver:

$$\mathbf{q}_{i,j} = \mathsf{J}_i \mathbf{K}_j + \mathbf{t}_i \quad \text{with} \quad \mathsf{J}_i \stackrel{\text{def}}{=} \mathsf{M}_i \mathsf{E}^{-1} \quad \text{and} \quad \mathbf{K}_j \stackrel{\text{def}}{=} \mathsf{E}\mathbf{S}_j. \tag{5.21}$$

We call J_i an implicit camera matrix and $\mathbf{S}_j$ are implicit shape basis. Matrix E represents a corrective or an upgrading transform that maps the implicit to the explicit LRSM. The main difference between the implicit and the explicit LRSM is that J_i does not have to comply with the replicated block structure characterizing M_i. While the latter is at least trilinear, the former is bilinear.

Fitting this bilinear driver to point tracks has two main uses in practice. The first one is to clear out some errors from the tracks and to glue splitted tracks together. This can be achieved reliably if other priors on the scene structure are considered, as we review in §5.3.3. The second use is to consider this 2D driver as a first step towards recovering the explicit 3D driver in a stratified manner, as described in §5.3.4.1.

5.3.3 More Priors

It has been shown by several authors that using more priors is necessary to make the LRSM driver well-constrained (Aanæs and Kahl, 2002; Del Bue et al., 2006; Torresani et al., 2007). One reason is that the LRSM is empirical and thus very sensitive to the number of shape bases. It is usually overestimated in order to minimize the goodness of fit which results in overfitting and bad conditioning. Using priors allows the model to better constrain the extra shape bases. We review some generic priors, where we use generic to mean that they are not object specific.

A very simple prior, used in (Del Bue et al., 2006) is to assume that a part of the scene is rigid. Smoothness of the shape deformation is used in (Aanæs and Kahl, 2002), while (Torresani et al., 2007) uses a Gaussian distribution prior on the configuration weights, and a Linear Dynamics model to enforce camera smoothness. We propose two priors in §7.1.3. The first one enforces temporal smoothness, in terms of camera path and scene deformation. It thus acts on both the configuration weights and the camera matrix, and can be used directly in the implicit driver on the implicit camera matrices J_i. The second prior is a continuous surface prior. It is based on the assumption that points consistently close in the images must have closely spaced shape bases. These priors dramatically improve the generalization ability of the implicit LRSM driver. Other priors can be found in the literature, such as the inextensibility of a continuous surface (Salzmann et al., 2007b), or the temporally local priors optimized with Second-Order Cone Programming in (Salzmann et al., 2007a).

5.3.4 Low-Rank Structure-from-Motion

We review different methods for estimating the LRSM drivers so as to achieve Low-Rank SfM, and hence Monocular Deformable SfM in a generic manner. Most of the current methods consider the reprojection error as a criterion to measure the goodness of fit. They generally differ in how this is minimized.

Low-Rank SfM can be seen as an extension of the Tomasi-Kanade rigid factorization method (Tomasi and Kanade, 1992). Equation (5.21) can indeed by rewritten as the factorization of a data matrix:

$$\begin{pmatrix} \mathbf{q}_{1,1} & \cdots & \mathbf{q}_{1,m} \\ \vdots & \ddots & \vdots \\ \mathbf{q}_{n,1} & \cdots & \mathbf{q}_{n,m} \end{pmatrix} = \begin{pmatrix} \mathsf{J}_1 \\ \vdots \\ \mathsf{J}_n \end{pmatrix} \begin{pmatrix} \mathbf{K}_1 & \cdots & \mathbf{K}_m \end{pmatrix}. \tag{5.22}$$

There have been many other extensions of this method, for instance to the perspective camera model. Extensions that relate to Low-Rank SfM include multibody factorization (Costeira and Kanade, 1998), temporal factorization (Zelnik-Manor and Irani, 2004) and articulated chain recovery (Yan and Pollefeys, 2008).

5.3.4.1 The Stratified Approach

Most of the algorithms follows a stratified approach to reconstructing the explicit LRSM which was initially proposed in (Bregler et al., 2000). They usually use three steps:

1. **Factoring.** The data matrix is built and factored to get the bilinear implicit LRSM parameters J_i and $\mathbf{K}_j$ following equation (5.22). Most of the work assumes that the rank is known, that there is no missing and erroneous data. This problem can be solved using the SVD to get the closest rank r approximation to the data matrix (see §5.5 for more details). This situation however is not very realistic. We tackle the general problem in §7.1.2, as one of the most difficult instances of the matrix factorization problem. We consider cases where the data has missing and erroneous elements, the rank is unknown, and the model is empirical, meaning that there is no ideal rank to describe the data. We proposed a specialized version of our general matrix factorization framework of §5.5. We have showen in §7.1.1 how the implicit LRSM driver can be computed from point and curve correspondences, along with a TPS warp registering the different images in a video. The process can be stopped at this step if only the implicit LRSM driver is sought (with possible prior enforcement and nonlinear refinement as we show in §7.1.3).

2. **Upgrading.** The corrective transform E is estimated so that the replicated block structure in $\mathsf{M}_i = \mathsf{J}_i\mathsf{E}$ holds as well as possible. This is a difficult step. The most recent solutions are reported in (Brand, 2005; Xiao and Kanade, 2006).

3. **Refining.** This step is optional. The reprojection error is minimized over all model parameters in a bundle adjustment manner. This is ill-posed in the sense that any shape basis can be replace by a linear combination of the other ones without changing the predicted points. The algorithms use different penalties to regularize the solution, as for instance the distance between two 3D contiguous shapes in the image sequence (Aanæs and Kahl, 2002).

The upgrading step requires the implicit camera matrices J_i to be split in triplets of columns, so as to enforce the block structure in the explicit camera matrices M_i. This means that the rank r must be estimated at step 1 as a multiple of 3, and the shape bases all are 3D. It actually turns out that there might be shape bases with smaller dimensions, *i.e.* 1D or 2D. This is studied in (Yan and Pollefeys, 2008).

5.3.4.2 The Probabilistic PCA Approach

The Probabilistic PCA approach has recently been proposed in (Torresani et al., 2007). They underline that the explicit LRSM is very sensitive to the number of shape bases, and often fails to reconstruct sensible cameras and shape bases. The problem is that there does not exist an ideal number of shape bases. They thus claim that priors, *i.e.* more priors than the LRSM, are needed so as to make the model well-behaved, and propose to use a Gaussian distribution prior on the configuration weights, in a Probabilistic PCA manner. This allows them to marginalize the configuration weights out of the estimation, which can then be performed very efficiently. They also propose to model temporal camera smoothness through a Linear Dynamics models with a transition matrix estimated jointly with the other parameters.

5.3.4.3 The Coarse-to-Fine Approach

We have proposed the coarse-to-fine approach in §7.1.4. It is motivated and inspired by the Deformotion paper (Yezzi and Soatto, 2003). The first ingredient is to define the motion and the deformation of a non-rigid object. This is what the Deformotion framework brings, by defining the motion with respect to some group of transformations, as the transformations and mean shape which best 'explain' the data. Based on this definition, we use rigid SfM to compute the mean shape and the camera parameters. This also is how (Aanæs and Kahl, 2002) initialize their reconstruction process. The second ingredient is that each shape basis should encapsulate as much as possible of the data variance remained unexplained by the coarsest shape bases, which is in accordance with the principle of PCA. We thus estimate the shape bases in turn by minimizing some cost including the reprojection error and some priors, leading to an efficient, coarse-to-fine algorithm that turns out to be very stable.

5.3.5 Selecting the Number of Shape Bases

As we mention above and is reported in several papers, choosing the right number of shape bases is crucial for the LRSM drivers to perform well. This becomes less important when priors are used, and can be overestimated, but still has to be determined. Obviously, increasing the number of shape bases makes the driver more flexible. This tough problem does not fit in the classical model selection framework. Indeed, classical model selection criteria such as AIC, BIC, GRIC and MDL often are derived in closed-form based on the fact that the prediction to data residuals follow some parametric distribution, often a Gaussian (see *e.g.* (Kanatani, 1998; Torr, 2002)). With the empirical LRSM, this is not likely to happen, since the residuals are mainly due to how the driver deviates from the physics, and are marginaly influenced by the noise on the point positions. We note that (Yan and Pollefeys, 2006) propose a method based on inspecting the eigenvalues of the data matrix.

We have proposed using Cross-Validation to select the number of shape bases in §§7.1.1 and 7.1.4. This technique does not require that the residuals follow a known parametric distribution. Details on Cross-Validation are given in §5.4. The idea is to partition the data in a training and a test set, and average the

test error over several such partitions. This approach, which has rarely been used for geometric model selection in computer vision, does not require a specific known distribution of the residuals, and directly reflects the ability of the model to extrapolate to new data. More precisely, we use v-fold Cross-Validation, which splits the data into v subsets or 'folds'. Typical values for v range from 3 to 10. We split the data image-point-wise, *i.e.* each fold is a subset of image points. The test error is obtained by comparing the test dataset with its prediction.

The typical behaviour of the Cross-Validation score is to decrease until the optimal number of shape bases is reached, and then to increase. It first decreases since without enough bases, the model is too restrictive to explain the data well, and thus cannot make good predictions. It then increases, since with more shape bases than strictly required, the model fits unwanted effects in the data, *i.e.* it is too flexible to only predict new instances. This typical behaviour however is not what we observe when the priors are used. In this case, the Cross-Validation score decreases rapidly until the optimal number of shape bases is reached, and then remains steady. This is explained by the fact that the priors diminish the degrees of freedom of the extra shape bases, as also reported in (Torresani et al., 2007).

5.3.6 Extensions

There are many possible extensions to the multilinear drivers. One of them would be to use kernels.[15] A way to choose a kernel would be to combine a Feature-Driven 2D warp as described in the previous section with a multilinear driver. The result is that the nonlinearly lifted point coordinates in (5.4) can be generated by a multilinear model. Provided that the warp is derived in a Reproducing Kernel Hilbert Space (Wahba, 1990), we expect the lifting function to play the role of the kernel, and the resulting driver to be a Kernel PCA-like one (Schölkopf et al., 1998). We note that Kernel PCA has been used for single-view fitting of an active shape model in (Romdhani et al., 1999).

5.4 The Prediction Sum of Squares Statistic and Cross-Validation

The results in this section are mostly related to the following papers:

[J11]	(§6.2.2)	*Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation*
[I17]	(§7.1.1)	*Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences*
[I50]	(§7.1.4)	*Coarse-to-Fine Low-Rank Structure-from-Motion*
	(§9.2.1)	*On Computing the Prediction Sum of Squares Statistic in Linear Least Squares Problems with Multiple Parameter or Measurement Sets*
[N15]	(§9.2.2)	*Reconstruction de surface par validation croisée*

I have directy applied greedy Leave-One-Out Cross-Validation during my Post Doctoral fellowship with Andrew Zisserman at the University of Oxford to select the number of shape bases in non-rigid factorization [I17]. I have used it similarly in the coarse-to-fine Low-Rank SfM method proposed in [I50] with my PhD student Vincent Gay-Bellile and a number of colleagues. I have shown how Leave-One-Out Cross-Validation can be used to estimate 2D warps in [J11] and have proposed new non-iterative formulaes with application to warp estimation. Finally, we have used Cross-Validation in [N15] with my PhD student Florent Brunet and his co-supervisors Nassir Navab from the Technical University of Munich and Rémy Malgouyres from Université d'Auvergne.

Choosing between multiple motion models is necessary to handle the so-called degenerate cases in rigid SfM, as is reported in *e.g.* (Pollefeys et al., 2002; Torr, 2002). This is typically done by testing several candidate models in a model selection framework. A related problem occurs when dealing with deformable models: it often is the case that the model to be estimated has a varying complexity, that can be tuned either by changing the number of model parameters or by using a smoother in the cost function. This is commonly done in *ad hoc* manners or by trial and error. A common example of this is the one of setting the smoothing parameter, which weights the smoother in compound data term plus smoother cost functions. Often, this is performed by visually inspecting the result. More sophisticated techniques are given in (Fua, 2000; Fua and Leclerc, 1995). The cost

[15] This idea comes from discussions with Andrew Zisserman during my Post Doctoral fellowship at the University of Oxford.

function can obviously not be minimized over the smoothing parameter since the result would always be zero, in other words, the most general model would always win.

Most of the problems we have tackled can be seen as Machine Learning problems. We have borrowed a technique that allows tuning the model complexity: the PRESS introduced in (Allen, 1974), which is very close to the Leave-One-Out Cross-Validation score (LOOCV) from (Wahba and Wold, 1975). It is related to the Jackknife and bootstrap techniques of sampling the dataset so that statistics can be drawn from it, and has been widely applied in Machine Learning (see *e.g.* (Bishop, 1995)). These techniques have marginally been used in computer vision for the reason that they are seen as requiring computationally intensive processing. This actually is partly true: there exist a non-iterative formula for the PRESS for LLS problems. We have shown that a similar formula gives a very good approximation to the LOOCV score, and that it can be extended to many different forms of non standard LLS problems. The key idea underlying these techniques is to make the model as general as possible in the sense of making it able to generalize to new data. This is different from the classical approach that makes the model fit the data as best as possible, given some fixed complexity. This is strongly inspired by the Machine Learning paradigm of supervised learning from examples, for which the classical approach is called Empirical Risk Minimization.

A successful approach is to consider the expected prediction error, also termed the *test error* or the *generalization error*, which, as the model complexity varies, measures the bias-variance trade-off (see *e.g.* (Poggio et al., 2004)). For the problems we are interested in, the generalization error can not be computed exactly since the number of data is usually low and their distribution is unknown. There are other ways to approximate the generalization error. The so-called model selection criteria such as BIC, AIC and GRIC have been successfully applied to pick the best model in a discrete set of possible models. For instance, given two images of a rigid scene, one must choose between, say a homography and the fundamental matrix (Kanatani, 1998; Torr, 2002). Determining the complexity of a deformable model does not fit in this setting since most of the models we use are empirical, making unlikely a possible parameterization of the model prediction-to-data residuals. A related approach is MDL, that has been used in medical image registration to register sets of multiple images (Marsland et al., 2008), and for SfM (Maybank and Sturm, 1999).

We point out that Cross-Validation is very different from the RANSAC paradigm (Fischler and Bolles, 1981). The latter trains the model using randomly sampled sets of minimal data, test on the rest of the data, and keeps the model with the largest 'consensus set'. It is meant to robustly estimate the model parameters, while Cross-Validation aims at quantifying the predictivity of the model. It is not obvious how RANSAC could be used to estimate image warps since there is not a clear definition of what a minimal dataset is in this case. However, Cross-Validation is not robust, in the sense that it does not cope with mismatched landmarks.

We have used Cross-Validation in a number of problems. In all these problems, an empirical model with varying complexity has to be fitted:

- ▷ **Implicit non-rigid factorization: selecting the rank.** When factoring the data matrix, its rank has to be determined. We have shown in §7.1.1 that Cross-Validation gives sensible results.

- ▷ **Coarse-to-fine Low-Rank SfM: selecting the number of shape bases.** In our algorithm reported in §7.1.4, shape bases are added until the Cross-Validation score increases or stabilizes.

- ▷ **Surface model fitting: selecting the smoothing weight.** We demonstrate an efficient algorithm that selects the smoothing weight for fitting an FFD surface in §9.2.2.

- ▷ **Warp estimation: selecting the smoothing weight and the number of control points.** We show how the LOOCV can be used to select the smoothing weight and the PRESS can be used to select the number of control points in §§6.2.2 and 9.2.1 respectively.

Organization of this section. First, we give general points about the PRESS statistic and the LOOCV score. Second, we report existing non-iterative formulas for computing these.

5.4.1 General Idea

The PRESS and LOOCV criteria are very similar in spirit. The difference is that the former is for cost functions with only a data term while the latter is for compound cost functions with a data term and smoothers. The basic idea is to measure the predictivity of a model with respect to a certain smoothing parameter, if any, by splitting the data into a training and a test set.

5.4.1.1 The Prediction Sum of Squares Statistic

The PRESS is typically used to compare different models. Let $\mathbf{u}$ be the parameter vector, f the model and $a_j \leftrightarrow b_j$ be m data points. Consider a regular NLS problem with the cost function:

$$\mathcal{E}^2_{\text{NLS}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}) - b_j\right)^2.$$

Fitting the model without the j-th measurement, gives the parameter vector $\mathbf{u}^\star_{\text{NLS},(j)}$:

$$\mathbf{u}^\star_{\text{NLS},(j)} \stackrel{\text{def}}{=} \arg\min_{\mathbf{u}} \frac{1}{m-1}\sum_{j'=1,j'\neq j}^{m}\left(f(a_{j'};\mathbf{u}) - b_{j'}\right)^2.$$

Note that the normalizing factor $\frac{1}{m-1}$ could be dropped without changing the solution. This is used to predict the j-th measurement as $f(a_j;\mathbf{u}^\star_{\text{NLS},(j)})$. This prediction is compared against the actual measurement b_j. This is averaged over the m measurements, giving the PRESS:

$$\mathcal{K}^2_{\text{NLS}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}^\star_{\text{NLS},(j)}) - b_j\right)^2. \tag{5.23}$$

This can typically be used to assess the predictivity of a warp against the number of control points. Note that this formula does not directly apply for the case of homogeneous problems (*i.e.* with vanishing right hand side).

5.4.1.2 The Leave-One-Out Cross-Validation Score

The LOOCV score is typically used to choose smoothing parameters. We now consider a regularized NLS problem with the cost function:

$$\mathcal{E}^2_{\text{RNLS}}(\mathbf{u};\mu) \stackrel{\text{def}}{=} \mathcal{E}^2_{\text{NLS}}(\mathbf{u}) + \mu^2\mathcal{E}^2_s(\mathbf{u}),$$

where $\mathcal{E}_s$ is the regularizer or smoother, for example $\|\mathbf{u}\|^2_2$ for ridge regression. The LOOCV score is defined similarly to the PRESS, but is a function of the smoothing parameter μ. We define:

$$\mathbf{u}^\star_{\text{RNLS},(j)}(\mu) \stackrel{\text{def}}{=} \arg\min_{\mathbf{u}} \frac{1}{m-1}\sum_{j'=1,j'\neq j}^{m}\left(f(a_{j'};\mathbf{u}) - b_{j'}\right)^2 + \mu^2\mathcal{E}^2_s(\mathbf{u}).$$

Contrarily to the regular NLS case, the normalizing factor $\frac{1}{m-1}$ can not be dropped. The LOOCV score is given by:

$$\mathcal{G}^2_{\text{RNLS}}(\mu) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(f(a_j;\mathbf{u}^\star_{\text{RNLS},(j)}(\mu)) - b_j\right)^2. \tag{5.24}$$

This can typically be used to assess the predictivity of a warp against the smoothing parameter.

A common way to speed up the computation is to use v-fold Cross-Validation. The idea is to partition the data into v folds with, typically, $v = 3,\ldots,10$. Each fold then serves as a test set in turn while the model is trained on the $v-1$ remaining ones. We use this approximation for Low-Rank SfM in §7.1.4.

5.4.2 Non-Iterative Solutions for Regular Linear Least Squares

The above formulas (5.23) and (5.24) for the PRESS and LOOCV are in general quite computationally expensive since they require solving the problem as many times as the number of data. There fortunately exist non-iterative formulas for LLS problems (Gentle et al., 2004). For instance, Tarpey (Tarpey, 2000) examines the case of restricted LLS. The PRESS formulas generally give the exact statistic value. The LOOCV formulas however are generally approximations to the true score, as we show in §6.2.2.

5.4.2.1 Deriving Non-Iterative Formulaes

We consider a regular LLS problem. Let A be the design matrix with m rows $\mathbf{a}_j$, $j = 1, \ldots, m$ and $\mathbf{b}$ the m measurement vector. The cost function is:

$$\mathcal{E}^2_{\text{STD}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{a}_j^{\mathsf{T}}\mathbf{u} - b_j\right)^2 = \frac{1}{m}\|\mathsf{A}\mathbf{u} - \mathbf{b}\|_2^2.$$

The solution to this problem is:

$$\mathbf{u}^{\star}_{\text{STD}} \stackrel{\text{def}}{=} \mathsf{A}^{\dagger}\mathbf{b}.$$

The PRESS is defined by fitting the model without the j-th measurement, giving the parameter vector $\mathbf{u}^{\star}_{\text{STD},(j)}$. This is used to predict the j-th measurement as $\mathbf{a}_j^{\mathsf{T}}\mathbf{u}^{\star}_{\text{STD},(j)}$. This prediction is compared against the actual measurement b_j. This is averaged over the m measurements, giving:

$$\mathcal{K}^2_{\text{STD}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{a}_j^{\mathsf{T}}\mathbf{u}^{\star}_{\text{STD},(j)} - b_j\right)^2.$$

Directly using this formula for estimating the PRESS would be extremely inefficient since the model has to be fitted m times to compute all the $\mathbf{u}^{\star}_{\text{STD},(j)}$. However, it is well-known that there is a non-iterative formula giving the PRESS as:

$$\mathcal{K}^2_{\text{STD}} = \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1} - \text{diag}\left(\hat{\mathsf{A}}\right)}\right)\left(\hat{\mathsf{A}} - \mathrm{I}\right)\mathbf{b}\right\|_2^2, \tag{5.25}$$

with $\hat{\mathsf{A}} = \mathsf{A}\mathsf{A}^{\dagger}$ the hat matrix. Note that $\left(\hat{\mathsf{A}} - \mathrm{I}\right)\mathbf{b} = \mathsf{A}\mathbf{u}^{\star}_{\text{STD}} - \mathbf{b}$, *i.e.* it is the residual vector. Formula (5.25) is proved in for example (Montgomery and Peck, 1992). It is equivalent to the sum of studentized residuals.

This has been used to find a non-iterative solution to compute the LOOCV score. We assume that the smoother has also an LLS form:

$$\mathcal{E}^2_s(\mathbf{u}) \stackrel{\text{def}}{=} \|\mathsf{Z}\mathbf{u}\|_2^2,$$

for some matrix Z. The hat matrix $\hat{\mathsf{A}}$ is replaced by the *influence matrix* defined as:

$$\mathsf{T}(\mu) \stackrel{\text{def}}{=} \mathsf{A}\left(\mathsf{A}^{\mathsf{T}}\mathsf{A} + m\mu^2\mathsf{Z}^{\mathsf{T}}\mathsf{Z}\right)^{-1}\mathsf{A}^{\mathsf{T}},$$

and the LOOCV score is approximated by:

$$\mathcal{G}^2_{\text{RSTD}}(\mu) \approx \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1} - \text{diag}\left(\mathsf{T}(\mu)\right)}\right)\left(\hat{\mathsf{A}} - \mathrm{I}\right)\mathbf{b}\right\|_2^2, \tag{5.26}$$

We demonstrate in §6.2.2 that this is a very good approximation to the true LOOCV score. The approximation comes from the m factor in the influence matrix.

Formula (5.26) leads to another approximation called Generalized Cross-Validation (Wahba, 1990), based on using $\text{diag}(\mathsf{T}(\mu)) \approx \text{tr}(\mathsf{T}(\mu))\mathrm{I}$ that allows one to simplify some calculations. We extend the formulaes (5.25) and (5.26) to non standard LLS cases in §§6.2.2 and 9.2.1. These include multiple linked parameter and measurement sets and is used for warp estimation purposes.

5.4.2.2 Minimizing the Statistics

The strategy for the PRESS is generally to start from a low number of parameters, and gradually add new ones until the statistic starts to increase, as we do in §6.2.2. The Cross-Validation score $\mathcal{G}_{\text{RSTD}}$ is a function of the smoothing parameter μ. We thus have to solve $\min_\mu \mathcal{G}^2_{\text{RSTD}}(\mu)$, which is a 1D minimization problem. A typical strategy is to draw sample smoothing parameters and select the one with the smallest Cross-Validation score, with optionaly some local polynomial interpolation of the score (see *e.g.* (Golub and von Matt, 1997; Hawkins and Yin, 2002)). For those cases where we have a non-iterative formula such as (5.26), other strategies are possible. The Cross-Validation score usually has a convex shape when plotted against the smoothing parameter, though this is not guaranteed. Possible minimization strategies include Golden Section Search (Burrage et al., 1994) and gradient descent, with $\mu_0 = 0$ as an initial solution. We used downhill simplex which we found in §6.2.2 to be the fastest method for the warp estimation problem.

5.5 Matrix Factorization with Missing and Erroneous Data

The results in this section are mostly related to the following papers:

[I22]	(§7.1.2)	*A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery*
[J10]	(§7.1.3)	*Implicit Non-Rigid Structure-from-Motion with Priors*
[I33]	(§8.1.1)	*Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion*

I have started working on the problem of matrix factorization in the context of SfM with Nicolas Guilbert and Anders Heyden from the University of Lund in 2002. We have published a first paper on how the closure constraints can be used with an affine camera [I13], and have extended it to a journal version where we have shown how perspective cameras can be initialized from affine ones [J07]. I have then started using the method for deformable SfM with Søren Olsen from the University of Copenhagen in 2005. We published our results in [I22], and have shown how priors can be incorporated in [I42,J10]. I have recently worked on this topic with Jean-Philippe Tardif from the University of Montréal in 2006. We have proposed the basis constraints applied to SfM in [I33]. Our latest experiments apply our algorithms to SfM, photometric stereo, non-rigid factorization and collaborative filtering.

Several computer vision and Machine Learning problems can be formulated as the one of matrix factorization, which is finding a rank r matrix $\mathsf{M}^\star$ as close as possible to a given data matrix M. This problem arises in *e.g.* linear dimensionality reduction, PCA, collaborative filtering (Goldberg et al., 2001), SfM for rigid scenes (Sturm and Triggs, 1996; Tomasi and Kanade, 1992) and deformable scenes (Bregler et al., 2000), illumination based reconstruction (Hayakawa, 1994), motion segmentation (Vidal and Hartley, 2004) and separation of style and content (Tenenbaum and Freeman, 2000).

For a fully available, *i.e.* complete data matrix with noisy inlying data, this is solved by the SVD (see *e.g.* (Srebro and Jaakkola, 2003)) as in the seminal affine SfM by factorization paper (Tomasi and Kanade, 1992). However, missing and erroneous data are unavoidable in many real-life situations. This makes the factorization problem much more complicated since it cannot be immediately solved with the SVD method anymore.

This section presents the methods we have proposed to solve the matrix factorization problem in the presence of missing and erroneous elements. We use complete blocks of the data matrix to compute constraints on one of the two factors. They make possible to estimate it using LLS. The other factor can then also be estimated using LLS. Using an analogy with SfM, estimating the first factor is computing the camera motion, while estimating the second one is finding the scene structure by triangulation, which is an affine LLS problem when using the affine camera model.

The constraints we have used are the *closure constraint*, which were proposed in (Triggs, 1997b) in the context of rigid perspective SfM, and the *basis constraint* that we have proposed as a dual to the closure constraint. We have described batch algorithms using the closure constraint, the basis constraint and a combination of those. The latter solution is in general the most stable one.

The great advantages of these algorithms is that the whole process is performed within seconds of computation using convex optimization routines. This includes the important early step of finding complete blocks

in the data matrix. The solutions given by these algorithms are so close to the global minimum that alternation schemes such as (Hartley and Schaffalitzky, 2003; Lu et al., 1997) almost always subsequently find the global minimum, in contrast to sequential approaches or random initialization procedures. Efficient NLS algorithms based on damped Newton are used in (Buchanan and Fitzgibbon, 2005). They require minutes or hours of computation since they need to be combined with multiple random starting points so as to find the global minimum.

Organization of this section. First, we state the problem and review some previous work. Second, we derive the first-factor closure constraint of Triggs and the second-factor closure constraint. Third, we propose the first- and second-factor basis constraints. All the algorithms using these constraints are based on analyzing complete blocks from the measurement matrix. We propose such a block finding algorithm, and finally show how to deal with erroneous matrix elements. Details are given in §8.1.1.

5.5.1 Problem Statement and Some Previous Work

The basic case: a complete inlying data matrix. Let M be the $(n \times m)$ measurement matrix. The factorization problem is re-stated as the one of finding two factors A and B, which are matrices with size $(n \times r)$ and $(r \times m)$ respectively, by solving:

$$\min_{\mathsf{A},\mathsf{B}} \|\mathsf{M} - \mathsf{AB}\|_{\mathcal{F}}^2 .$$

This *factorization residual* is for instance proportional to the reprojection error in affine SfM (Reid and Murray, 1996). The Frobenius norm is used since we assume that M is an *i.i.d.* Gaussian noise corrupted rank r matrix:[16]

$$\mathsf{M} \stackrel{\text{def}}{=} \mathcal{M} + N(0, \sigma^2) \qquad \text{with} \qquad \text{rank}(\mathcal{M}) = r \qquad \text{and} \qquad \mathcal{M} = \mathcal{A}\mathcal{B},$$

where $\mathcal{M}$ is the noise-free matrix and $\mathcal{A}$ and $\mathcal{B}$ its two noise-free factors. This is an NLS problem, which has a simple solution in practice given by computing the SVD of matrix M:

$$\mathsf{M} \stackrel{\text{SVD}}{\longrightarrow} \mathsf{U}\Sigma\mathsf{V}^{\mathsf{T}},$$

and taking, *e.g.*:

$$\mathsf{A} \leftarrow \bar{\mathsf{U}}\sqrt{\Sigma'} \qquad \text{and} \qquad \mathsf{B} \leftarrow \sqrt{\Sigma'}\bar{\mathsf{V}}^{\mathsf{T}},$$

where $\bar{\mathsf{U}}$ and $\bar{\mathsf{V}}$ contain the r leading columns of U and V, and Σ' contains only the r leading singular values of M taken from Σ. We note that this solution is not unique since there is a gauge freedom ambiguity. Let C be an $(r \times r)$ full-rank matrix, then the factors AC and $\mathsf{C}^{-1}\mathsf{B}$ give a solution equivalent to A and B.

Missing data: a partial inlying data matrix. The problem gets more complicated when some entries of the data matrix are missing. Given the binary $(n \times m)$ missing data indicator matrix W (called the visibility matrix in the context of SfM), the problem is stated as:

$$\min_{\mathsf{A},\mathsf{B}} \|\mathsf{W} \odot (\mathsf{M} - \mathsf{AB})\|_{\mathcal{F}}^2 .$$

Many different methods have been proposed, most of them dedicated to SfM, which, depending on the camera model, can be formulated as rank-3 matrix factorization with a translational part. Broadly speaking, the methods can be classified as *iterative* and *batch*. The SVD based technique is batch since it uses all the data almost equally. For the missing data case, iterative techniques are the most popular. Directly applying an NLS Newton-based optimization algorithm has been attempted in (Buchanan and Fitzgibbon, 2005) with random starting points. We believe that this kind of methods should only be the final step, and should proceed from an initial estimate lying as close as possible to the global minimum.

[16] Matrix factorization is also used for problems where the data matrix is empirically supposed to be rank r, *e.g.* collaborative filtering or non-rigid SfM, see §7.1.3. Some problems require other noise models, leading to different cost functions.

Missing and erroneous data: a partial data matrix with outliers. The problem is even more difficult when some of the data can be erroneous. Using an element-wise matrix M-estimator ρ, one possible problem statement is:

$$\min_{\mathsf{A},\mathsf{B}} \|\rho\left(\mathsf{W} \odot (\mathsf{M} - \mathsf{A}\mathsf{B})\right)\|_{\mathcal{F}}^2 .$$

This cost function can be minimized by Newton-based techniques. The algorithms we provide below are made robust using RANSAC, as we describe in §5.5.7.

5.5.2 Overview of our Batch Algorithms and Application to Structure-from-Motion

The different batch algorithms we have proposed are all based on the following main steps. They start by selecting a number of complete blocks from the data matrix, from which constraints on one of the factors are formed. Once the factor has been solved for, the other one is computed using standard, possibly robustified, affine LLS. The optional final step is to refine both factors together by minimizing the factorization residual using damped Newton iterative NLS. Thereafter, we assume for simplicity of writing that the first factor is estimated first.

Our batch algorithms can be applied to SfM as described in details in §8.1.1. If an affine camera model is used, then it consists of rank-4 factorization of a data matrix made with image point coordinates with a translational part that has to be explicitly dealt with. For the perspective camera model, the projective depth of each image point has to be recovered first in order to rescale the data matrix made of homogeneous point coordinates. This can be done using one of the techniques in (Martinec and Pajdla, 2005a; Sturm and Triggs, 1996).

5.5.3 The Closure Constraints and Estimation Algorithms

The closure constraints from (Triggs, 1997b) allow one to estimate the first factor, *i.e.* matrix A, without estimating the second factor, *i.e.* matrix B. We reformulate these constraints into what we call *first-factor closure constraints*.

5.5.3.1 Deriving the Constraints

The idea is to consider a complete measurement block[17] $\tilde{\mathcal{M}}$ from the data matrix, to factor it using the SVD solution, and show that this gives constraints on the first factor $\mathcal{A}$. The block $\tilde{\mathcal{M}}$ is obtained by selecting a subset of rows in $\mathcal{M}$ by multiplying to the left by some row-amputated identity matrix Π, and a subset of columns by multiplying to the right by some column-amputated identity matrix Γ:

$$\tilde{\mathcal{M}} \stackrel{\text{def}}{=} \Pi\mathcal{M}\Gamma. \tag{5.27}$$

We choose blocks with rank at least r. Matrix $\tilde{\mathcal{M}}$ can be factored using an SVD to give:

$$\tilde{\mathcal{M}} \stackrel{\text{SVD}}{\longrightarrow} \mathsf{U}\Sigma\mathsf{V}^{\mathsf{T}}, \tag{5.28}$$

Let the size of the selected block be $(\tilde{n} \times \tilde{m})$. The r leading columns of U form a basis for $\tilde{\mathcal{M}}$, while the remaining $\tilde{m} - r$ columns $\tilde{\mathcal{N}}$ of U form a basis for the left kernel of $\tilde{\mathcal{M}}$, *i.e.* we have $\tilde{\mathcal{N}}^{\mathsf{T}}\tilde{\mathcal{M}} = 0$. For noise corrupted data, $\tilde{\mathcal{N}}$ is the best approximation to the left kernel of $\tilde{\mathsf{M}}$ since it minimizes $\|\tilde{\mathcal{N}}^{\mathsf{T}}\tilde{\mathsf{M}}\|_{\mathcal{F}}$. Substituting $\tilde{\mathcal{M}}$ by its definition (5.27) and $\mathcal{M}$ by its factorization $\mathcal{M} = \mathcal{A}\mathcal{B}$ gives:

$$\tilde{\mathcal{N}}^{\mathsf{T}}\Pi\mathcal{A}\mathcal{B}\Gamma \;=\; 0.$$

Both $\Pi\mathcal{A}$ and $\mathcal{B}\Gamma$ are rank r at most. Any element in the vector space spanned by the columns $\tilde{\mathcal{N}}$ thus lies in the left kernel of $\Pi\mathcal{A}$. This gives the *first-factor closure constraint*:

$$\mathcal{N}^{\mathsf{T}}\mathcal{A} \;=\; 0 \qquad \text{with} \qquad \mathcal{N}^{\mathsf{T}} \stackrel{\text{def}}{=} \tilde{\mathcal{N}}^{\mathsf{T}}\Pi,$$

[17] $\tilde{\mathcal{M}}$ is the noise free version of $\tilde{\mathsf{M}}$.

where $\mathcal{N}$ has been dubbed *matching tensor* on the original SfM work of Triggs, since it relates to the multiple view matching tensors such as the fundamental matrix. The sparsity of $\mathcal{N}$ is directly related to the block size. In practice, choosing small to medium size blocks ensures that the design matrix in the global system is highly sparse.

The *second-factor closure constraint* is obtained by examining the right kernel of $\tilde{\mathcal{M}}$, or equivalently, by replacing $\mathcal{M}$ by $\mathcal{M}^{\mathsf{T}}$ in the above derivation. This gives constraints on the second factor.

5.5.3.2 Estimating the First Factor

The closure constraints directly lead to LLS algorithms for estimating A. Let $\mathcal{N}_1, \ldots, \mathcal{N}_l$ be l matchings tensors estimated for different, noisy measurement blocks. We find the solution to the following homogeneous LLS problem:

$$\min_{\mathsf{A}} \|\mathsf{N}\mathsf{A}\|_{\mathcal{F}}^2 \qquad \text{s.t.} \qquad \text{rank}(\mathsf{A}) = r \qquad \text{with} \qquad \mathsf{N}^{\mathsf{T}} \stackrel{\text{def}}{=} (\mathcal{N}_1 \ \cdots \ \mathcal{N}_l). \tag{5.29}$$

Two ways can be used for solving this problem. The first one directly solves it under its homogeneous LLS form and takes advantage of the high sparsity of the design matrix by using Implicitly Restarted Arnoldi Methods (Arnoldi, 1951; Lehoucq and Scott, 1996). The second method transforms the problem (5.29) to an affine LLS one by fixing the gauge, *i.e.* by fixing an $(r \times r)$ full rank block of A to some arbitrary full rank matrix such as the identity matrix. This slightly changes the cost that is being minimized but leads to very close, slightly more accurate results.

5.5.4 The Basis Constraints

5.5.4.1 Deriving the Constraints

The idea of our basis constraints comes from the block factorization (5.28). Closure constraints only use the left kernel of $\tilde{\mathcal{M}}$, given by the $\tilde{m} - r$ last columns of U in the SVD, but ignore its orthonormal r leading columns, that we denote by $\bar{\mathsf{U}}$. These columns form a basis of $\Pi\mathcal{A}$. Hence, there exists an $(r \times r)$ full rank *alignment matrix* Z such that:

$$\Pi\mathcal{A} \ = \ \bar{\mathsf{U}}\mathsf{Z}.$$

We call this equation the *first-factor basis constraints*. These constraints are in a sense dual to the closure constraints since they form a generating basis of the unknowns, as opposed to direct constraints. In SfM, they correspond to computing a partial reconstruction expressed in its own coordinate frame.

5.5.4.2 Estimating the First Factor

The basis constraints directly lead to LLS algorithms for estimating A. Let $\bar{\mathsf{U}}_1, \ldots, \bar{\mathsf{U}}_l$ be l bases obtained from the l different measurement blocks given by the $\Pi_1, \ldots, \Pi_l$ row selecting matrices. Solving for A requires one to also solve for the transformations $\mathsf{Z}_1, \ldots, \mathsf{Z}_l$, bringing all the bases in the same coordinate system. In practice, this amounts to solving:

$$\min_{\mathsf{A},\mathsf{Z}_1,\ldots,\mathsf{Z}_l} \sum_{k=1}^{l} \|\bar{\mathsf{U}}_k\mathsf{Z}_k - \Pi_k\mathsf{A}\|_{\mathcal{F}}^2 \qquad \text{s.t.} \qquad \text{rank}(\mathsf{A}) = r. \tag{5.30}$$

This is rewritten with a single matrix norm as:

$$\min_{\mathsf{A},\mathsf{Z}_1,\ldots,\mathsf{Z}_l} \left\| \begin{pmatrix} \Pi_1 & -\bar{\mathsf{U}}_1 & & 0 \\ \vdots & & \ddots & \\ \Pi_l & 0 & & -\bar{\mathsf{U}}_l \end{pmatrix} \begin{pmatrix} \mathsf{A} \\ \mathsf{Z}_1 \\ \vdots \\ \mathsf{Z}_l \end{pmatrix} \right\|_{\mathcal{F}}^2 \qquad \text{s.t.} \qquad \text{rank}(\mathsf{A}) = r.$$

Fixing the gauge leads to an affine LLS problem with a block arrowhead shape design matrix, which frequently appears, for instance in Orthogonal Distance Regression problems (Boggs et al., 1989). This makes it possible

to solve problem (5.30) without constructing its design matrix explicitly. As shown previously for the case of the closure constraints, the homogeneous LLS system can also be directly solved. In SfM, this method is analogous to a one-level hierarchical approach of computing partial reconstructions, and registering them altogether at once, as is done in *e.g.* (Fitzgibbon and Zisserman, 1998; Martinec and Pajdla, 2005b).

5.5.5 Combining Closure and Basis Constraints

The two types of constraints, *i.e.* closure and basis, are equivalent from an algebraic point of view. They however give different results in practice since they involve minimizing different cost functions to find the first factor. It is thus natural to try combining them together. This can be done easily since both (5.29) and (5.30) are homogeneous LLS problems:

$$\min_{\mathsf{A},\mathsf{Z}_1,\dots,\mathsf{Z}_l} \left\| \begin{pmatrix} \Pi_1 & -\bar{\mathsf{U}}_1 & & 0 \\ \vdots & & \ddots & \\ \Pi_l & 0 & & -\bar{\mathsf{U}}_l \\ \tilde{\mathcal{N}}_1^\mathsf{T}\Pi_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \tilde{\mathcal{N}}_l^\mathsf{T}\Pi_l & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathsf{A} \\ \mathsf{Z}_1 \\ \vdots \\ \mathsf{Z}_l \end{pmatrix} \right\|_{\mathcal{F}}^2 .$$

This is an LLS problem of the same type as (5.30), so it can be solved efficiently in a similar manner.

Gathering the two sets of constraints together is obviously arbitrary in the sense that they are of different nature, so at least a parameter should somehow balance them. We however found in our experiments in §8.1.1 that this straightforward joint use of the constraints lead to an improvement of the results.

5.5.6 Finding Complete Blocks

Finding complete blocks in the data matrix is a key step in our algorithms. One has to find sufficiently many, carefully chosen blocks, so that one of the factors can be retrieved using the closure or the basis constraints. The primary constraint that the set of blocks have to fulfill is that each row of the data matrix has to be involved in at least r blocks. If the second factor were to be computed first, each column of the data matrix would have to be implied in at least r blocks. We have to take into account whether the constraints are robustly estimated from each block. For instance, estimating a first-factor closure constraint using RANSAC requires that the block has a sufficient number of columns to allow the point-wise random sampling step to be efficient. The number of columns does not change the computational cost required to combine the constraints. We note that finding the largest complete block is an NP-hard problem (Jacobs, 2001).

Our algorithm takes as inputs the data and the missing element matrices M and W, the rank r and the minimum number of constraints d to be associated to each row.[18] We distribute the constraints as evenly as possible amongst the rows. The idea is to sequentially scan the columns of the data matrix. Recalling that the rows and columns forming a block do not have to be contiguous, we randomly select between $r+1$ and $3r$ rows and check whether these rows contain enough non empty columns to create a block. The $3r$ limit ensures that the algorithm can build blocks easily. To make the search fast and to evenly distribute the constraints among the rows, we count the constraints associated to each row, and consider the rows with smallest count to build the next block. Blocks are added until all rows have a sufficient number of constraints. The chances are low that a block is selected twice and in practice, the effect is negligible.

5.5.7 Dealing with Erroneous Data

Our batch algorithms basically consist of two rounds of convex optimization for estimating each of the two factors in turn. They can handle erroneous data if both steps are robustified, as follows. The first step starts

[18] It turns out that associating each row with at least $d > r$ constraints improves the results compared to using the minimal $d = r$, see the experimental results in §8.1.1 for more details.

by computing constraints from blocks taken from the data matrix. These blocks usually involve a redundent number of columns, and can be robustly computed while selecting the inlying columns using RANSAC, and checking that the block is non degenerate. This is typically what is done in SfM for computing multiple view matching tensors. The first step then estimates the first factor by combining all the constraints together. One could think of making this step robust as well, by *e.g.* using RANSAC or an M-estimator with Iteratively Reweighted Least Squares over the different constraints. This in practice is not useful since robustified constraint computation procedures usually successfully reject the outlying columns in each block. This however produces a very rough inlier/outlier classification of the data since a whole block column is either kept or rejected, while we would like each data to be given a label. This can be achieved by robustifying the second step. Indeed, assuming that the first factor has been correctly retrieved, the second one is computed column-wise. This makes it possible, for each column of the data matrix, to apply RANSAC so as to select the inlying rows, thereby producing an element wise labeling of the data matrix. This is typically what triangulation algorithms do in SfM. We have applied this scheme with great success in rigid SfM, see 8.1.1, and non-rigid SfM, see §§7.1.2 and 7.1.3.

5.6 Compositional and Learning-Based Image Registration

The results in this section are mostly related to the following papers:

[I40]	(§6.1.1)	*Direct Image Registration With Gain and Bias*
[J12]	(§6.1.2)	*Groupwise Geometric and Photometric Direct Image Registration*
[V01]	(§6.2.4)	*Feature-Driven Direct Non-Rigid Image Registration*
[I29]	(§7.1.5)	*Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model*

I have started working on learning-based registration with my PhD students Vincent Gay-Bellile and Mathieu Perriollat in 2006. We have published a method for deformable surface tracking in [I29]. We have then extended the method in several ways, in particular with the use of a compositional framework for non-groupwise warps, and a piecewise linear prediction model in [V01]. I have proposed a method that allows estimating both geometric and photometric parameters within a compositional framework. I have used a trick specific to gain and bias [I40], and have then proposed a more general method based on the so-called photometric inverse compositional rule [J12].

Image registration is commonly done by minimizing an NLS cost function that can be feature- or pixel-based, as described in §5.2. Other solutions are possible, such as the one in (Glocker et al., 2007), combining Markov Random Fields with linear programming or those using Graph Cuts as for instance in (Boykov and Kolmogorov, 2004). Iterative methods such as Gauss-Newton and Levenberg-Marquardt with theoretical superlinear convergence have shown to be very effective, given a decent initial solution. Examples include the Lucas-Kanade algorithm (Lucas and Kanade, 1981) and bundle adjustment techniques (Triggs et al., 2000). These methods are based on a local linearization of the cost function, leading to so-called normal equations, that give the update vector at each iteration. Both the design matrix and the right hand side vector in the normal equations vary through the iterations. They thus have to be recomputed, and the system has to be fully solved at each iteration..

It has been shown in (Baker and Matthews, 2004) that under some hypotheses, the design matrix can be made constant. In other words, only the right hand side vector has to be recomputed at each iteration, while the design matrix can be precomputed and the system 'presolved',[19] thereby saving a significant load of computation. This is made possible by using a compositional parameter update rule, as we briefly explain below.

Let $\mathbf{u} \in \mathbb{R}^p$ be the set of parameters to be estimated, for example for a image warp $\mathcal{W}$ as in (5.1), *i.e.* such that $\mathbf{q}' = \mathcal{W}(\mathbf{q}; \mathbf{u})$. An additive parameter update rule is implicitly used in the above mentioned classical algorithms:

$$\mathbf{u} \leftarrow \mathbf{u} + \boldsymbol{\delta}, \tag{5.31}$$

[19] In practice, matrix factorization techniques are used to solve to system in a stable manner.

where $\boldsymbol{\delta}$ is the update vector computed at each optimization iteration. This can be equivalently written as:

$$\mathcal{W}(\cdot;\mathbf{u}) \leftarrow \mathcal{W}(\cdot;\mathbf{u}+\boldsymbol{\delta}).$$

The compositional algorithms are based on a compositional parameter update rule. The geometric forward compositional update rule is defined as:

$$\mathcal{W}(\cdot;\mathbf{u}) \leftarrow \mathcal{W}(\mathcal{W}(\cdot;\boldsymbol{\delta});\mathbf{u}),$$

and the geometric inverse compositional update rule is defined as:

$$\mathcal{W}(\cdot;\mathbf{u}) \leftarrow \mathcal{W}(\mathcal{W}^{-1}(\cdot;\boldsymbol{\delta});\mathbf{u}). \tag{5.32}$$

Why these rules are called 'geometric' will be made clear shortly. Using a compositional update rule has been first proposed in (Shum and Szeliski, 2000) who showed that compositional update rules usually lead to simpler Jacobian matrix. The forward compositional update rule requires that the warp can be composed, while the inverse compositional one also requires that the warp can be inverted. In other words, they respectively require that the warp has a semi-group and a group structure. This is the case for many global warps such as the affine transformation or the homography, but does not generally hold for deformable warps such as FFD and RBF warps. Thereafter, we assume that a pixel-based cost function such as (5.3) is used. Though compositional update rules can be used with feature-based cost functions, see *e.g.* (Benhimane and Malis, 2007), deriving a constant design matrix iteration would take different steps from the pixel-based case.

Organization of this section. First we give general points about compositional image registration algorithm. Second, we show how inverse composition can be performed jointly on a geometric and a photometric transformation. Third, we review existing work on how to deal with non-groupwise transformations. Finally, we review means for the local forward registration step based on piecewise linear motion prediction.

5.6.1 General Points

We give a derivation of the inverse compositional image registration algorithm for estimating a geometric transformation, *i.e.* a warp, with parameter vector $\mathbf{u}_g$. Starting from the pixel-based cost function (5.3), we have to solve the following problem:

$$\min_{\mathbf{u}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))\|_2^2.$$

We introduce the inverse compositional update rule (5.32) on $\mathbf{u}_g$. The optimization is now to be over $\boldsymbol{\delta}_g$, leading to:

$$\min_{\boldsymbol{\delta}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathcal{W}^{-1}(\mathbf{q};\boldsymbol{\delta}_g);\mathbf{u}_g))\|_2^2.$$

The first approximation to be made is to apply the incremental transformation to the source image, instead of the target one. The is called the inverse compositional trick, and leads to:

$$\min_{\boldsymbol{\delta}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathcal{W}(\mathbf{q};\boldsymbol{\delta}_g)) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))\|_2^2.$$

Note that this is an approximation since the cost function is now expressed within the coordinate system of the warped image and not within the source image as the original one.

The second approximation to be made is to use the Gauss-Newton approximation over the update parameter vector. Assuming that $\mathcal{W}(\cdot;\mathbf{0})$ is the identity warp, this approximation is to be made around $\mathbf{0}$:

$$\min_{\boldsymbol{\delta}_g} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) + \mathsf{L}_g^\mathsf{T}(\mathbf{q})\boldsymbol{\delta}_g - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))\|_2^2,$$

where, using the chain rule, the Jacobian matrix L_g is given by:

$$\mathsf{L}_g^{\mathsf{T}}(\mathbf{q}) \stackrel{\text{def}}{=} (\nabla\mathcal{S})(\mathbf{q})^{\mathsf{T}}(\nabla_{\mathbf{u}_g}\mathcal{W})(\mathbf{q};\mathbf{0}).$$

This is an LLS problem for $\boldsymbol{\delta}_g$. The entries of the design matrix thus depend on the L_g matrices, which are independent of the parameter vector. The design matrix, and thus the normal equations can be 'presolved', so that the update parameter vector is given by multiplying some right hand side vector by a constant, precomputed matrix. The elements of the right hand side vector are $\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))$, in other words, the pixel color of the difference or residual image, since $\mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g))$ are the pixel color of the warped image.

This paradigm, proposed in (Baker and Matthews, 2004), has been shown to be very efficient in terms of computational cost for an iteration, and in terms of convergence properties (it has a large convergence basin, and quickly reaches the sought after solution).

Summing up, the Gauss-Newton inverse compositional algorithm precomputes the local registration design matrix and 'presolves' the normal equation, and then iterates the following three main steps:

- ▷ **Step 1: Warping.** Warp the target image towards the source one using the current warp parameters $\mathbf{u}_g$ and compute the difference image. This is done by combining the warp with a simple, *e.g.* bilinear, image interpolation scheme.

- ▷ **Step 2: Local registration.** Register the warped target to the source image to get the update parameter vector $\boldsymbol{\delta}_g$. This is *e.g.* done using LLS through Gauss-Newton approximation of the cost function, or using a learning approach as described in §5.6.4.

- ▷ **Step 3: Updating.** Update the parameter vector $\mathbf{u}_g$ using the inverse compositional update rule (5.32). This is straightforward if the warp has a group structure, but requires a special procedure if not, as described in §5.6.3.

Convergence is typically determined by comparing the norm of the update vector to some threshold such as 10^{-4}. Note that other approximations to the cost function can be used for local registration, such as Efficient Second-order Minimization (ESM) (Benhimane and Malis, 2007). The next section shows how this algorithm can be extended to jointly deal with a photometric transformation.

5.6.2 Geometric and Photometric Inverse Composition

We propose the dual inverse composition algorithm in §6.1.2. Its purpose is to efficiently compute both the geometric and a photometric registration. Other algorithms which estimate photometric registration such as those in (Baker et al., 2003) usually either fully solve the normal equations at each iteration, or use approximations that spoil the convergence frequency and may dramatically increase the number of iterations. Our algorithm uses the inverse compositional update trick for both the geometric and photometric counterparts of the registration, thereby making it possible to 'presolve' the normal equations.

We introduce a photometric transformation $\mathcal{V}$ to be applied to pixel colors with parameter $\mathbf{u}_p$. We show that the inverse composition update rule for such transformations differs from (5.32) since the inverse incremental transformation must be composed 'to the left' of the current one. We thus define the photometric inverse compositional update rule as:

$$\mathcal{V}(\cdot;\mathbf{u}_p) \leftarrow \mathcal{V}^{-1}(\mathcal{V}(\cdot;\mathbf{u}_p);\boldsymbol{\delta}_p). \tag{5.33}$$

The registration problem is stated as the one of solving:

$$\min_{\mathbf{u}_g,\mathbf{u}_p} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{V}(\mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g));\mathbf{u}_p)\|_2^2.$$

Combining the geometric and photometric inverse compositional update rules (5.32) and (5.33) we get the dual inverse compositional update rule:

$$\mathcal{V}(\mathcal{T}(\mathcal{W}(\mathbf{q};\mathbf{u}_g));\mathbf{u}_p) \leftarrow \mathcal{V}^{-1}(\mathcal{V}(\mathcal{T}(\mathcal{W}(\mathcal{W}^{-1}(\mathbf{q};\boldsymbol{\delta}_g);\mathbf{u}_g));\mathbf{u}_p);\boldsymbol{\delta}_p). \tag{5.34}$$

The problem is thus rewritten as:

$$\min_{\boldsymbol{\delta}_g,\boldsymbol{\delta}_p} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{V}^{-1}(\mathcal{V}(\mathcal{T}(\mathcal{W}(\mathcal{W}^{-1}(\mathbf{q};\boldsymbol{\delta}_g);\mathbf{u}_g));\mathbf{u}_p);\boldsymbol{\delta}_p)\|_2^2.$$

Using several approximations leading to a cost function expressed in the geometric and photometric coordinate frames of the warped image, and the Gauss-Newton approximation on both $\boldsymbol{\delta}_g$ and $\boldsymbol{\delta}_p$, we obtain that the update parameter vectors can be computed by simply multiplying a constant matrix by some right hand side vector. This derivation uses the assumption that the geometric and photometric transformations commute, *i.e.* that $\mathcal{V}(\mathcal{T}(\mathcal{W}(\cdot;\mathbf{u}_g));\mathbf{u}_p) = (\mathcal{V}(\mathcal{T};\mathbf{u}_p))(\mathcal{W}(\cdot;\mathbf{u}_g))$. This assumption holds for any global photometric transformations such as affine ones. The algorithm can be applied to those photometric transformations that, similarly to the warp, have a group structure. An example of this is the gain and bias transformation accounting for global and uniform lighting change. Let $\mathbf{v}$ be a pixel color vector, we define:

$$\mathcal{V}(\mathbf{v};\mathbf{u}_p) \;=\; u_{p,1}\mathbf{v} + u_{p,2}. \tag{5.35}$$

The first element of $\mathbf{u}_p$ thus is the gain and the second one is the bias. We show how more complex transformations can be dealt with in §6.1.2, including a full, 12 parameter, affine transformation.

5.6.3 Handling Non-Groupwise Warps

The efficient inverse compositional algorithm requires that the transformation to be estimate has a group structure, *i.e.* that composing two transformations gives a transformation of the same kind, and that a transformation can be inverted. Most deformable image warps do not have such a group structure, preventing the use of compositional algorithms.

Approaches for non-groupwise warp composition are proposed in (Matthews and Baker, 2004; Romdhani and Vetter, 2003) in the context of fitting a face 3DMM. In this case, the parameter vector is the state of the 3DMM and the camera pose. They usually solve the problem in two steps. First, the previous model vertices are transferred to the current image by applying the local and then the global warp. They usually are not in accordance with a projected instance of the 3DMM. Second, the parameter update vector is recovered by minimizing a prediction error, namely the distance between the updated vertices and those predicted by the model. This last step requires nonlinear optimization, which may be expensive since it takes place in the inner loop of the main fitting procedure. Warp inversion is approximated with first order Taylor expansion in (Matthews and Baker, 2004), while triangular meshes are used in (Romdhani and Vetter, 2003), which thereby does not require linearization. These two methods are straightforward to adapt to the case of warp estimation.

The method we have proposed in §6.2.4 allows approximating the composition and inversion of deformable warps in closed-form. The backbone of this approach is the Feature-Driven warp parameterization, which naturally arises with FFD warps, and that we derived in §5.2.2.4 for RBF warps.

5.6.4 Learning-Based Local Registration

Departing from the usual paradigm of computing a local approximation to the cost function with linear or Gauss-Newton expansion, some authors suggested learning the local registration, *i.e.* the parameter update vector, as a function of the difference image. This idea dates back to (Cootes et al., 1998). It has been used in (Jurie and Dhome, 2002) for homography estimation, and more recently in (Matas et al., 2006). The relationship is learned offline from synthetically generated training data. It fits very well in the forward compositional framework, since it allows estimating the forward update parameter vector by simply multiplying one of multiple constant matrices with the right hand side vector containing the difference image.

Related learning approaches in the literature assume that the relationship between the error image and the update parameters is linear (Cootes et al., 1998; Jurie and Dhome, 2002; Matas et al., 2006). A single *interaction matrix* is thus learned. The drawback of these methods is that the motion scale, *i.e.* the average displacement magnitude, of the training data is difficult to choose to cover all cases. On the one hand, if the interaction matrix covers a large domain of deformation magnitudes, the alignment accuracy is spoiled. On the

other hand, if the matrix is learned for small deformations only, the converge basin is dramatically reduced. Interaction matrices are valid only locally around the texture image parameters. Compositional algorithms are thus required, as in (Jurie and Dhome, 2002) for homographic warps. The Feature-Driven framework naturally extends this approach to non-groupwise warps. However, (Cootes et al., 1998) makes the assumption that the domain where the linear relationship is valid covers the whole set of registrations. They thus apply the single interaction matrix around the current parameters, avoiding the warping and the composition steps. This does not appear to be a valid choice in practice.

The strategy in (Matas et al., 2006) is different and leads to a feature-based method. A set of keypoints is selected as the points that can be best used for linear prediction. This is then used along with RANSAC on a global motion model, namely a homography, to discard the outliers. In the context of deformable warps, it could be used with a robust warp estimation procedure such as the one in (Pilet et al., 2008).

We note that there are other work using Machine Learning for tracking as *e.g.* the seminal Singular Vector Tracking paper (Avidan, 2004) and (Williams et al., 2005). These techniques could possibly be used for image registration. The method we have proposed in §6.2.4 overcomes the above mentioned problems by using a piecewise linear relationship between the difference image and the local registration. A statistical test is trained so as to choose which linear part of the predictor should be used given the difference image.

CHAPTER

6

IMAGE REGISTRATION

In this chapter we study the 2D image registration problem. This is equivalent to the computation of a geometric warp matching two images.

In the first part we tackle the photometric issues occuring in pixel-based methods. These typically arise when the two images to be registered are under different illuminations. We have followed two approaches. In the first one, we estimate a global lighting change, that we explicitly model at the pixel color level. We have proposed methods for estimating parametric photometric transformations. These methods are formulated in the inverse compositional image registration framework thanks to our photometric inverse compositional update rule. In the second approach, we estimate a non-uniform lighting change, including cast shadows. Our strategy has been to project the image to a 1D light invariant space, into which we register the images. Global explicit transformations need to be estimated jointly so as to correct the pixel color prior by projecting it to the 1D light invariant space.

In the second part, we examine the deformable image registration problem. We have defined what we call the Generalized Thin-Plate Spline warps, that incorporate perspective projection effects and rigidity constraints to regular Thin-Plate Spline warps. We show how the smoothing parameter can be automatically estimated using Cross-Validation. Our other contributions concern several aspects of pixel-based deformable image registration. These include the automatic insertion of deformation centres, a framework for inverse compositional deformable image registration, learning-based local registration and the detection of self-occlusions.

6.1 Photometry in Pixel-Based Image Registration

This section is devoted to the work we have done for the modeling and estimation of photometric transformations between images. In the first part, we present two papers. They are inspired by (Baker et al., 2003) whose several algorithms are proposed to extend their inverse compositional image registration framework (Baker and Matthews, 2004) to deal with geometric warps and photometric transformations. The algorithms proposed in (Baker et al., 2003) are very general in that they deal with linear appearance variations. They however spoil the efficient inverse compositional framework by either recomputing the Jacobian matrix at each iteration or by approximating the original cost function. This makes these algorithms slow or unreliable. Our two papers propose algorithms that are more focused that those in (Baker et al., 2003). They are called the *gain and bias inverse compositional algorithm* and the *dual inverse compositional algorithm*. The former one handles affine photometric transformations with up to 12 parameters.

In the second part we present another paper. It draws on the light invariance image framework of (Finlayson et al., 2002). This framework allows one to 'project' a color image to a light invariant space. (Finlayson et al., 2002) use it for shadow removal purposes, as also demonstrated in (Finlayson et al., 2004). We propose to use it for *shadow resistant image registration*.

6.1.1 Paper (LIMA3D'06) – *Direct Image Registration With Gain and Bias*

V01 **Direct Image Registration With Gain and Bias**
A. Bartoli
Topics in Automatic 3D Modeling and Processing Workshop, Verona, Italy, March 2006

The main contribution of this paper is the *gain and bias inverse compositional algorithm*. We study the direct registration problem of two single channel images. The warp can be any groupwise transformation such as a homography and the photometric model is a 1D affine transformation. The problem statement is:

$$\min_{\mathbf{u}_g,\mathbf{u}_p} \sum_{\mathbf{q}\in\mathcal{P}} (u_{p,1}\mathcal{S}(\mathbf{q}) + u_{p,2} - \mathcal{T}(\mathcal{H}(\mathbf{q};\mathbf{u}_g)))^2.$$

Using an inverse compositional update rule (5.32) on the warp parameters $\mathbf{u}_g$ and an additive update rule (5.31) on the photometric parameters $\mathbf{u}_p$ does not lead to a constant Jacobian matrix. We however show that an efficient solution to the normal equations is achieved using a trick from photogrammetric block bundle adjustment. The Hessian matrix is pre-inverted blockwise at the off-line stage. The blocks are combined on-line with weights depending on the gain $u_{p,1}$. This gives the inverse Hessian and the solution to the normal equations. This approach is fast and has very good convergence properties. A shortcoming however is that it does not however extend to color images or to more complex photometric models.

6.1.2 Paper (PAMI'08) – *Groupwise Geometric and Photometric Direct Image Registration*

J12 **Groupwise Geometric and Photometric Direct Image Registration**
A. Bartoli
IEEE *Transactions on Pattern Analysis and Machine Intelligence*, accepted December 2007
Previous version: [I28]
Related paper: [I44]

The main contribution of this paper is the *dual inverse compositional algorithm*. It generalizes the above described gain and bias inverse compositional algorithm. It handles multiple channel images; the warp and the photometric transformation can be any groupwise transformations. In practice, we use an homographic warp $\mathcal{H}$. We tested various models for the photometric transformation $\mathcal{V}$, ranging from a simple 1D affine transformation to a full 12 parameter channel mixing affine transformation. The problem statement is given by:

$$\min_{\mathbf{u}_g,\mathbf{u}_p} \sum_{\mathbf{q}\in\mathcal{P}} \|\mathcal{S}(\mathbf{q}) - \mathcal{V}(\mathcal{T}(\mathcal{H}(\mathbf{q};\mathbf{u}_g));\mathbf{u}_p)\|_2^2.$$

Note that it is expressed within the source image, both geometrically and photometrically. In other words, the target image combined to the geometric warp and photometric transformation acts as a generator for the source image. Our algorithm is rooted in the photometric and dual inverse compositional rules (5.33) and (5.34) we have proposed, which makes the Hessian matrix constant. This allows us to precompute it and thus to 'presolve' the normal equations.

This approach is fast and has very good convergence properties. An example of this is shown in figure 6.1.

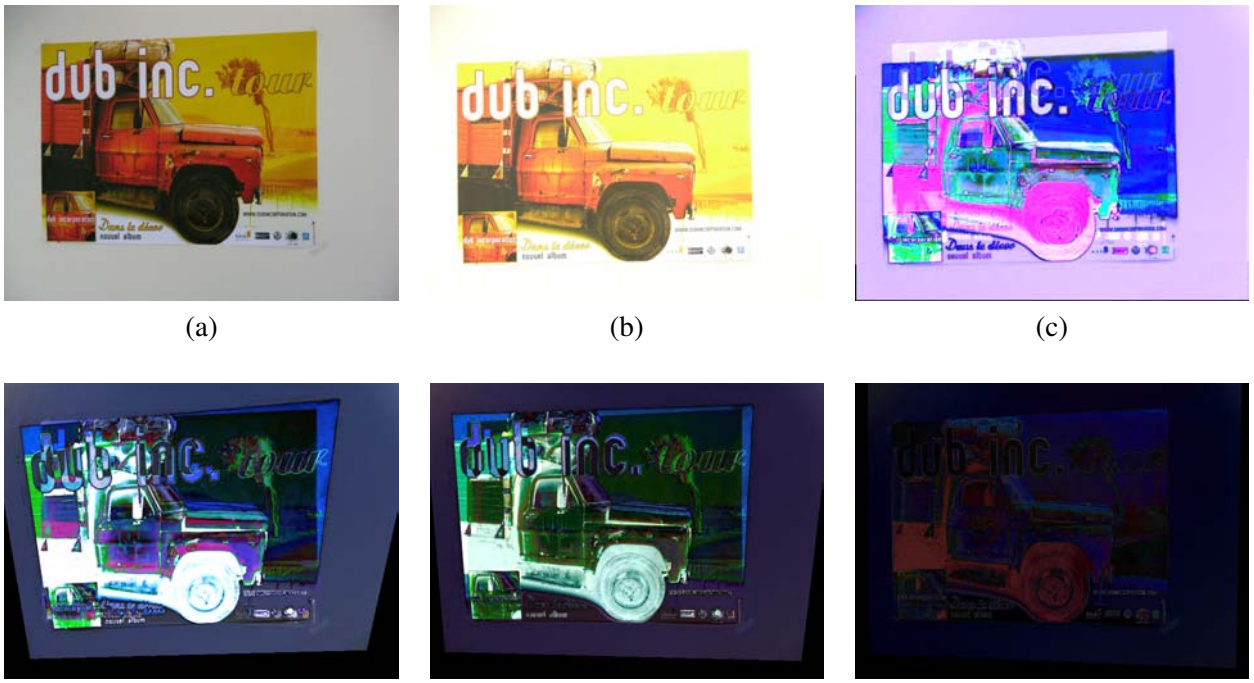


Figure 6.1: Paper (PAMI'08) – *Groupwise Geometric and Photometric Direct Image Registration*. Example of pixel-based geometric and photometric image registration. The two images were registered using our *dual inverse compositional algorithm*. (a) and (b) show the two original images. The lighting is different in intensity and color. (c) – (f) are a selection of difference images through the registration process, from the start to convergence.

6.1.3 Paper (SCIA'07) – *Shadow Resistant Direct Image Registration*

I31	**Shadow Resistant Direct Image Registration** D. Pizarro and A. Bartoli SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007

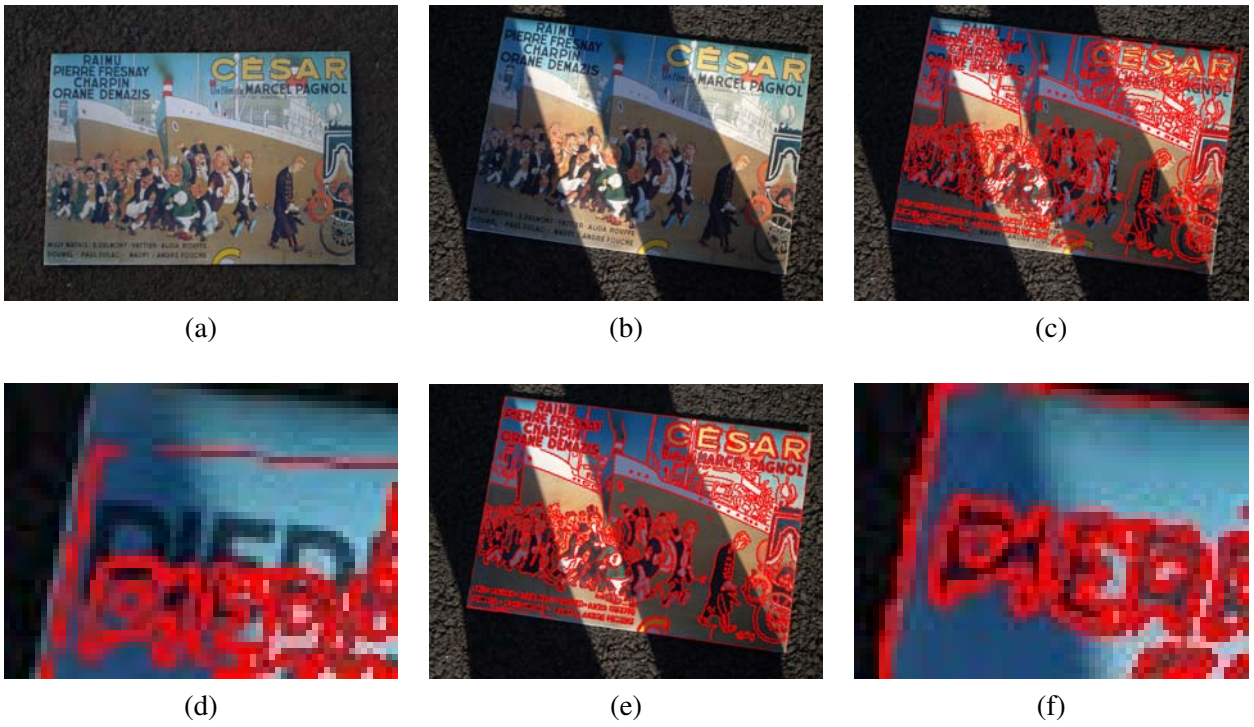
The main contribution of this paper is a *shadow resistant image registration algorithm*. It is based on the light invariance theory of (Finlayson et al., 2002). This theory is based on a physical model of the photometric camera response to a Lambertian surface illuminated by a Planckian light. It says that an RGB image can be 'projected' to a 1D light invariant image. This projection operator $\mathcal{L}$ takes RGB color vectors and a scalar parameter θ as inputs. This scalar parameter is related to the photometric response of the camera and must be estimated. (Finlayson et al., 2004) show how the θ parameter can be estimated from a single image containing shadows. A more detailed photometric model taking vignetting effects into account is used in (Kim and Pollefeys, 2008).

Our algorithm is based on minimizing the image difference in the light invariant space. This is different from the approaches which explicitly model the photometric transformation. Our approach requires us to find the θ parameter for each of the images, as well as some global photometric transformation. We experimentally

found that using a gain and bias transformation for each of the color channels performs well. The problem statement is:

$$\min_{\mathbf{u}_g,\mathbf{u}_p,\theta_s,\theta_t} \sum_{\mathbf{q}\in\mathcal{P}} (\mathcal{L}(\mathcal{S}(\mathbf{q});\theta_s) - \mathcal{L}(\mathcal{V}(\mathcal{T}(\mathcal{H}(\mathbf{q};\mathbf{u}_g));\mathbf{u}_p);\theta_t))^2.$$

Our algorithm solves for the geometric warp, and 'self-calibrates' the two cameras in a photometric manner. An example of this is shown in figure 6.2.



(a) (b) (c)

(d) (e) (f)

Figure 6.2: Paper (SCIA'07) – *Shadow Resistant Direct Image Registration*. Example of pixel-based image registration. (a) and (b) show the two original images. The lighting is different in intensity, color and shading. The results are illustrated by transferring and plotting in the target image contours extracted in the source one. (c) shows the result of a classical registration algorithm and (d) is a closeup on (c). (e) shows the result obtained by our algorithm and (f) is a closeup on (e).

6.2 Estimation of Deformable Image Warps

The section tackles the problem of estimating deformable image warps between two images. Both feature- and pixel-based cost functions are used. The first part has two papers centred on extending and estimating the TPS warp proposed in (Bookstein, 1989). The proposed extensions are perspective projection and scene rigidity. They mix up the TPS warp with visual geometry as described for instance in (Hartley and Zisserman, 2003). The estimation methods we propose are feature-based. They use the Prediction Sum of Squares (PRESS) (Allen, 1974) and Leave-One-Out Cross-Validation (LOOCV) (Wahba and Wold, 1975) techniques. These offer two different ways of selecting the complexity of a model. For some of the warps, the PRESS is defined in §9.2.1.

The second part brings three papers on the pixel-based estimation of deformable warps. Compositional update rules are not directly possible with deformable warp as reported in (Matthews and Baker, 2004). We propose a simple method for compositional and inverse compositional update of a deformable warp. We show that forward compositional image registration can be performed with learned piecewise linear prediction. Finally, we examine the problem of self-occlusions, that defeats most of the previous algorithms. We propose a solution based on a shrinker. This is a term which is added to the cost function that makes the warp shrinks along the self-occlusion boundary.

6.2.1 Paper (CVPR'07) – *Generalized Thin-Plate Spline Warps*

I34 **Generalized Thin-Plate Spline Warps**
A. Bartoli, M. Perriollat and S. Chambon
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

The main contributions of this paper are *several extensions to the well known TPS warp*. The basic idea is that the regular TPS warp is interpreted as being induced by a smooth deforming surface observed by a moving affine camera. We name this warp the *DA-Warp*, for Deformable Affine TPS Warp. We use our Feature-Driven parameterization of §5.2.2.4, based on the target centre coordinates in P'. We express everything in the lifted form (5.4) and its extension to perspective. With this parameterization, the DA-Warp is written:

$$\mathcal{W}_{\text{DA}}(\mathbf{q};\mathsf{P}',\lambda) \stackrel{\text{def}}{=} \mathcal{M}_{\text{DA}}\cdot\nu_{\text{TPS}}(\mathbf{q}) \qquad \text{with} \qquad \mathcal{M}_{\text{DA}}^{\mathsf{T}} \stackrel{\text{def}}{=} \mathsf{X}_\lambda\mathsf{P}'.$$

The first extension we propose is the *RA-Warp*, for Rigid Affine TPS Warp. It assumes that the observed surface is rigid. In other words, the warp must comply with the affine epipolar geometry. Our RA-Warp is parameterized by the affine fundamental matrix $\mathcal{A}$ and the depth of the centres in $\boldsymbol{\delta}$. It is written:

$$\mathcal{W}_{\text{RA}}(\mathbf{q};\boldsymbol{\delta},\mathcal{A},\lambda) \stackrel{\text{def}}{=} \mathcal{M}_{\text{RA}}\cdot\nu_{\text{TPS}}(\mathbf{q}) \qquad \text{with} \qquad \mathcal{M}_{\text{RA}}^{\mathsf{T}} \stackrel{\text{def}}{=} \mathsf{X}_\lambda(\mathsf{P}\ \boldsymbol{\delta}\ \mathbf{1})\mathcal{S}_{\mathcal{A}}^{\mathsf{T}},$$

with $\mathcal{S}_{\mathcal{A}}$ some (2×4) affine camera matrix associated to the target image in the canonical coordinate frame.

The second extension we propose is the *RP-Warp*, for Rigid Perspective TPS Warp. It assumes that the observed surface is rigid and observed by perspective cameras. Our RP-Warp is parameterized by the fundamental matrix $\mathcal{F}$ and the depth of the centres in $\boldsymbol{\delta}$. It is written:

$$\mathcal{W}_{\text{RP}}(\mathbf{q};\boldsymbol{\delta},\mathcal{F},\lambda) \stackrel{\text{def}}{=} \Psi\left(\mathcal{M}_{\text{RP}}\cdot\nu_{\text{TPS}}(\mathbf{q})\right) \qquad \text{with} \qquad \mathcal{M}_{\text{RP}}^{\mathsf{T}} \stackrel{\text{def}}{\sim} \mathsf{X}_\lambda(\mathsf{P}\ \boldsymbol{\delta}\ \mathbf{1})\mathcal{G}_{\mathcal{F}}^{\mathsf{T}},$$

with $\mathcal{G}_{\mathcal{F}}$ some (3×4) perspective camera matrix associated to the target camera in the canonical coordinate frame. This warp is in a lifted perspective form.

The third extension we propose is the *DP-Warp*, for Deformable Perspective TPS Warp. It assumes a deformable surface observed by perspective cameras. It is parameterized by the homogeneous coordinates of the target centres in $\tilde{\mathsf{P}}'$. It is written:

$$\mathcal{W}_{\text{DP}}(\mathbf{q};\tilde{\mathsf{P}}',\lambda) \stackrel{\text{def}}{=} \Psi\left(\mathcal{M}_{\text{DP}}\cdot\nu_{\text{TPS}}(\mathbf{q})\right) \qquad \text{with} \qquad \mathcal{M}_{\text{DP}}^{\mathsf{T}} \stackrel{\text{def}}{\sim} \mathsf{X}_\lambda\tilde{\mathsf{P}}'.$$

We show an example of this in figure 6.3. The way we constructed these Generalized TPS warps can be applied to other RBF warps in a straightforward manner. It could also be used with FFD warps, and would probably be strongly related to the Non Uniform Rational B-Splines (NURBS) to model perspective projection.

We defined a hierarchy between the four aforementioned warps. We showed that the set of DP-Warps contains all the other warps. The set of RA-Warps is the intersection of the set of DA-Warps and the set of RP-Warps. We also studied the asymptotic regularization behaviour of these warps. These warps are easily expressed thanks to the Feature-Driven parameterization and the lifted affine and perspective forms.

6.2.2 Paper (JMIV'08) – *Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation*

J11 **Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation**
A. Bartoli
Journal of Mathematical Imaging and Vision, special issue: tribute to Peter Johansen, accepted December 2007

In classical image registration, often a smoothing parameter is used which balances the data term and the smoother to form a compound cost function. This fixes the effective number of warp parameters. Choosing a 'good' smoothing parameter is very important in order to obtain sensible results. The main contribution of this paper is a method for *automatically choosing the smoothing parameter based on LOOCV*. More precisely, this

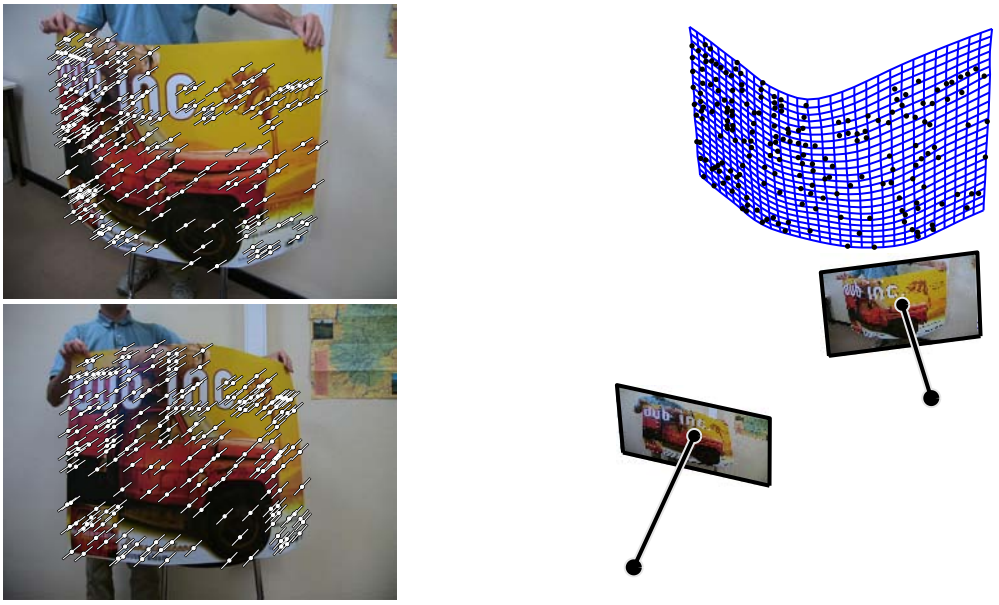
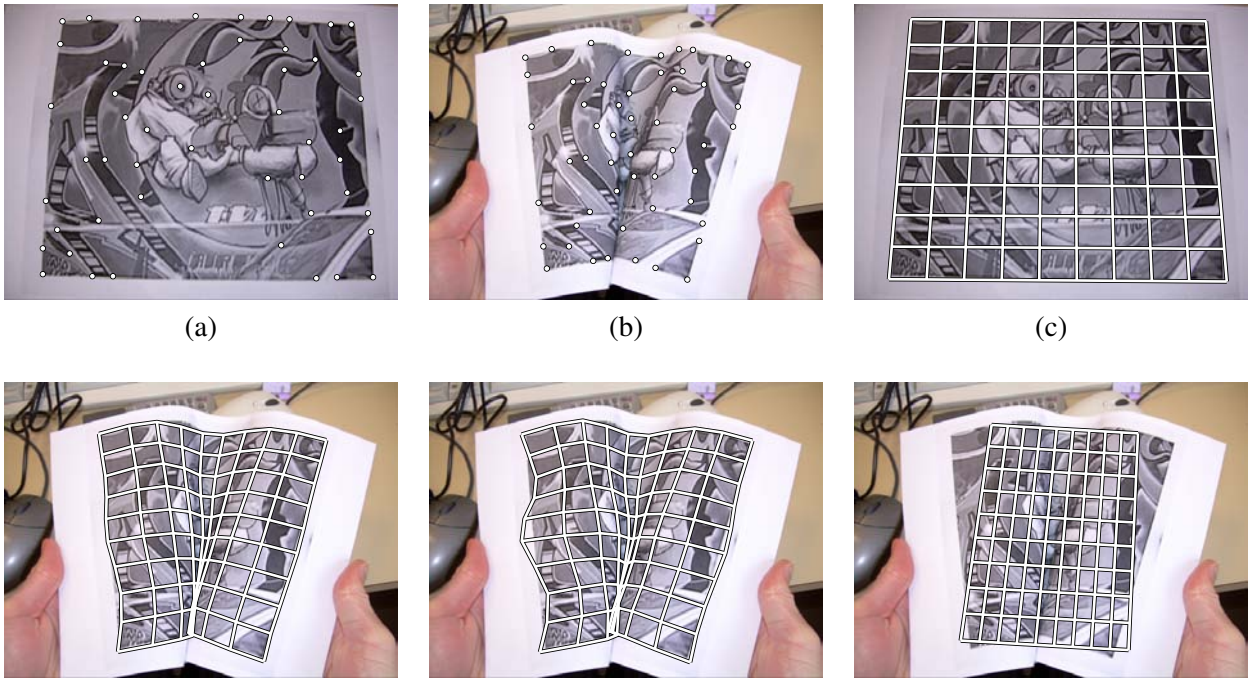


Figure 6.3: Paper (CVPR'07) – *Generalized Thin-Plate Spline Warps*. The two left images are 'perspective images' of a rigid smooth surface, overlaid with 206 point correspondences and epipolar lines. The right image shows the surface recovered through our RP-Warp.



(a) (b) (c)

(d) (e) (f)

Figure 6.4: Paper (JMIV'08) – *Maximizing the Predictivity of Smooth Deformable Image Warps Through Cross-Validation*. Example of automatic smoothing parameter selection with LOOCV. (a) and (b) show two images of a paper overlaid with 53 point correspondences. (c) shows a warp visualization grid manually set in the source image. (d) shows the warp visualization grid transferred to the target image with the automatically selected smoothing parameter. (e) and (f) respectively illustrate overfitting due to lack of smoothing, and oversmoothing.

is a feature-based method for the DA-Warp, *i.e.* the standard TPS warps, that also directly applies to any warp in the lifted affine form (5.4). These contributions are strongly linked to §5.4.

Our contributions are two-fold. We show that the well-known non-iterative formula (5.26) for LOOCV can be extended to deal with DA-Warps by simply replacing the vector two-norm by the matrix Frobenius norm. As described in (Wahba and Wold, 1975), this standard formula is commonly believed to give the true LOOCV score. We show that this formula is actually an approximation to the true LOOCV score. The reason might be that there exists a very similar formula which gives the PRESS with no approximation. We show that the non-iterative LOOCV formula is generally a good approximation, with its local minimum located very close to the minimum of the true LOOCV score. An example of this is shown in figure 6.4.

6.2.3 Paper (BMVC'04) – *Direct Estimation of Non-Rigid Registrations*

I18 **Direct Estimation of Non-Rigid Registrations**
A. Bartoli and A. Zisserman
BMVC'04 - *British Machine Vision Conf.*, London, UK, September 2004

The main contribution of this paper is a *pixel-based algorithm for estimating RBF warps with dynamical centre insertion*. The number of free parameters, *i.e.* the number of centres, is automatically chosen while estimating the warp. The key idea of the algorithm is to start with few centres and iteratively insert new centres as the registration proceeds. These new centres are chosen by inspecting the difference image. The pixel value in this image are interpreted as a measure of misregistration. Therefore, blobs in the difference image are used to give the location of new centres. Our algorithm estimates a simple gain and bias model (5.35) that accounts for global lighting change. Our algorithm takes the following steps, and terminates when the decrease in the norm of the difference image is not significant. First an affine transformation is estimated. Second a new centre is inserted, which position is given as follow. The source and the warped images are first blurred[1] and then the difference of these is taken. Then an integration step is performed by convolving the difference image with a Gaussian. The centre is inserted at the highest local maximum of the resulting image. Third, the algorithm minimizes the pixel-based error over all the warp parameters. It then loops back to the second step until convergence.

A second contribution we bring in this paper is a method for using pair-wise image registration to register a video. Without loss of generality, we choose one image of the video as the source image,[2] also called the 'reference image'. Registering the source image with each of its neighbours in the video can be achieved with *e.g.* the algorithm above. However, this kind of algorithms, based on the brightness constancy assumption, fails when the source and target images are too different in appearance from each other. This typically happens in videos, where *e.g.* shadows might appear, disappear or move on the surface. The solution we propose is to 'update' the source image. This is done by replacing the source image with the warped image obtained at convergence before proceeding to the next image in the video. This approach might be very effective but is prone to drifting since slight misregistrations accumulate while proceeding the video. It turns out that combining the original source image with the warped image gets rid of the drifting in most cases (this has been tested after publication of our paper). Another solution we have tried is to simply update a 'shadow mask', discarding those pixels which are detected in a shaded area in the previous image in the video.

6.2.4 Paper (BMVC'07) – *Feature-Driven Direct Non-Rigid Image Registration*

I40 **Feature-Driven Direct Non-Rigid Image Registration**
V. Gay-Bellile, A. Bartoli and P. Sayd
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007
Version in French: [N12]

[1] We do not blur the difference image directly to avoid effects such as the partial pixel effect.
[2] The source image could also be a template coarsely registrered to one of the video image.

The main contributions of this paper are *(i)* a *framework for using compositional algorithms with deformable image warps* and *(ii)* a *piecewise linear predictor for forward local registration*. These contributions are respectively related to §§5.6.3 and 5.6.4.

Threading and reversing warps. The backbone of this approach is the Feature-Driven warp representation we give in §5.2.2.4 for RBF warps. The idea is to parameterize the warp by a set of driving features. The coordinates of which thus form the parameter vector $\mathbf{u}_g$. Such as parameterization can also be derived for FFD warps. These features have a fixed position $\mathbf{r}_g$ in the source image, which depends on the type of warp that is being used. With this representation, a warp can be seen as an interpolant between the driving features. There is obviously an infinite number of such warps. The best one to use depends on the nature of the images. Loosely speaking, we say that matching the driving features between two images is equivalent to defining a warp since the warp can be used to transfer them from one image to the other while conversely, a warp can be computed from the driving features.

Our methods for warp composition and invertion are respectively called warp threading and reversion. They are based on very simple and intuitive ideas. For warp threading, we consider two sets of driving features, say $\mathbf{u}_g$ and $\mathbf{u}'_g$. We have to find a warp with driving features $\mathbf{u}''_g$ that behaves like the composition of the warps induced by $\mathbf{u}_g$ and $\mathbf{u}'_g$. Our idea is to apply the $\mathbf{u}'_g$ induced warp to the features in $\mathbf{u}_g$ to get $\mathbf{u}''_g$. For driving features lying well-spread in the region of interest, this gives very good results, and does not require optimization. Threading warp is thus simply done by using:

$$\mathbf{u}''_g \;=\; \mathcal{W}(\mathbf{u}_g;\mathbf{u}'_g).$$

Reversing a warp is also very simple. Considering a set of driving features in $\mathbf{u}_g$, we are looking for the driving features in $\tilde{\mathbf{u}}_g$ such that the warp they induce behaves like the inverse of the warp induced by $\mathbf{u}_g$. Our idea is that applying the $\tilde{\mathbf{u}}_g$ induced warp to $\mathbf{u}_g$ should give $\mathbf{r}_g$, *i.e.* the fixed driving features in the source image, which is written:

$$\mathcal{W}(\mathbf{u}_g;\tilde{\mathbf{u}}_g) \;=\; \mathbf{r}_g.$$

This gives an exactly determined, *i.e.* square, linear system whose solution is the sought after driving feature in $\tilde{\mathbf{u}}_g$.

Piecewise linear prediction. We propose to learn a piecewise linear relationship between the difference image and the update parameter vector. Concretely, we train a series of α interaction matrices $\mathcal{F}_1,\ldots,\mathcal{F}_\alpha$, each of which covers a different range of displacement magnitudes. A statistical matrix selection procedure is learned in order to select the most appropriate matrix $\mathcal{F}_\beta$ given the difference image $\mathcal{D}$, and the forward incremental parameter vector is simply given by $\boldsymbol{\delta}_g = \mathcal{F}_\beta \text{vect}(\mathcal{D})$.

An interaction matrix $\mathcal{F}$ is learned from artificially perturbed source images $\mathcal{A}_z$ with $z = 1,\ldots,t$. The driving features $\mathbf{r}_g$ in the source image are disturbed from their rest position with randomly chosen direction and magnitude to give $\mathbf{u}_{g,z} = \mathbf{r}_g + \boldsymbol{\delta}_{g,z}$. The latter is clamped between a lower and an upper bound, determining the area of validity of the interaction matrix to be learned. Our Feature-Driven warp reversion process is used to warp the source image. The difference image $\mathcal{D}_z$ is then computed, and the interaction matrix $\mathcal{F}$ learned by minimizing an LLS error in the image space, expressed in pixel color unit, as:

$$\mathcal{F} \;=\; \left(\mathcal{L}\mathcal{U}^\mathsf{T}\left(\mathcal{U}\mathcal{U}^\mathsf{T}\right)^{-1}\right)^\dagger \quad \text{with} \quad \mathcal{U} \stackrel{\text{def}}{=} (\boldsymbol{\delta}_{g,1} \;\cdots\; \boldsymbol{\delta}_{g,t}) \quad \text{and} \quad \mathcal{L} \stackrel{\text{def}}{=} (\mathcal{D}_1 \;\cdots\; \mathcal{D}_t).$$

This is one of the two possibilities for learning the interaction matrix. The other possibility is dual. It minimizes an error in the parameter space, *i.e.* expressed in pixels. Our experimental results show that the former approach performs much better. Note that we also used this approach in §7.1.5 so as to deal with a 3D warp, driven by a surface 3DMM. In this case, the training data are generated by perturbing, and then projecting, the 3D surface.

One issue with the piecewise linear model is to select the best interaction matrix at each iteration. Each of these indeed has a specific domain of validity in the displacement magnitude. Experimental results show that applying all the matrices in turn appears not to be the most discerning choice. The matrices for large displacements are applied first, which makes dramatically high the number of iterations needed to converge. This

shows the requirement of a sensible matrix selection criterion. The displacement magnitude can unfortunately not be determined prior to image alignment. We propose to learn a relationship between the magnitude of the difference image and the displacement magnitude intervals. We express this relationship with a Gaussian distribution. An example of this is shown in figure 6.5.

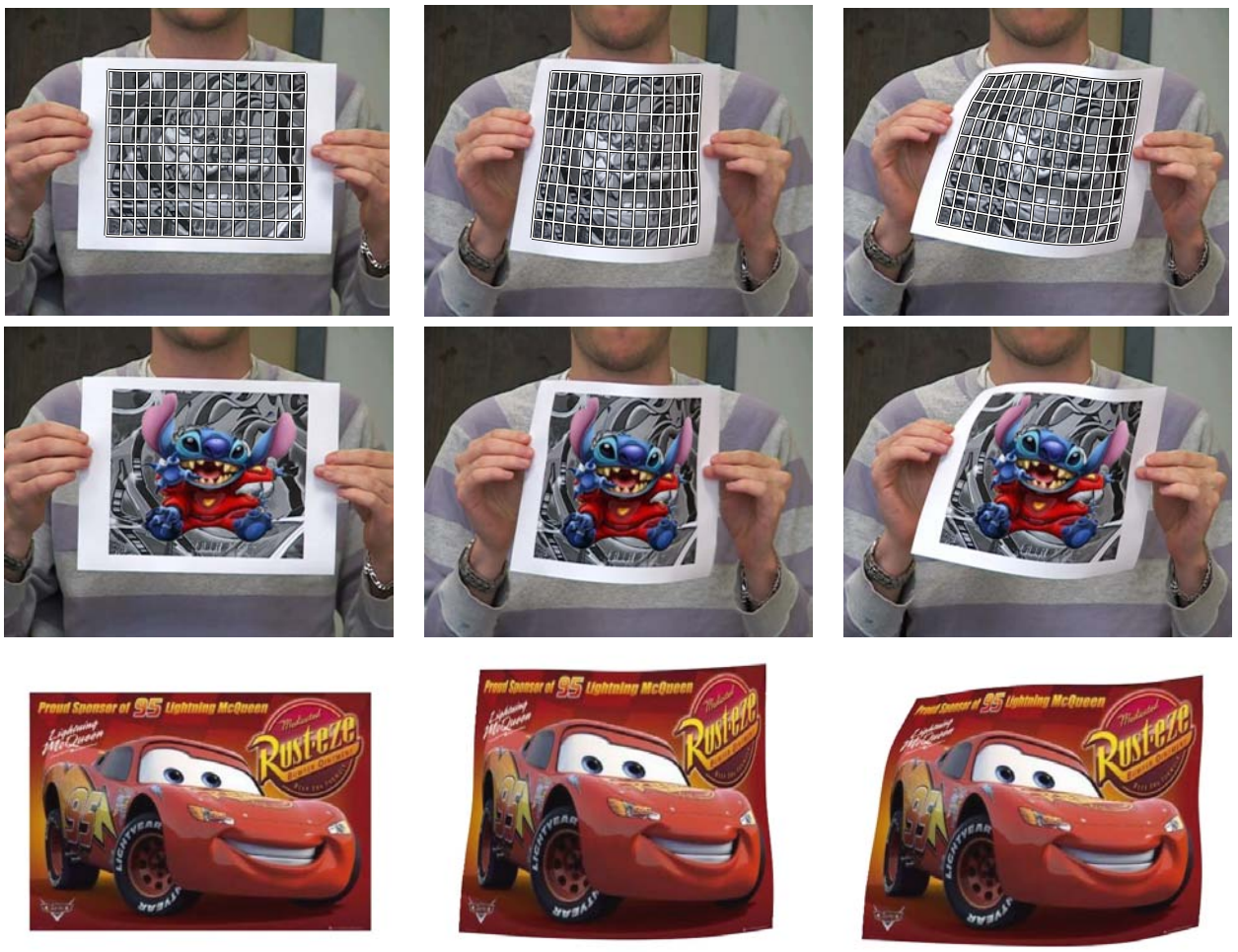


Figure 6.5: Paper (BMVC'07) – *Feature-Driven Direct Non-Rigid Image Registration*. Each column shows an image from a 350 image video of a deforming paper. The top row shows meshes illustrating the recovered warp. The middle row shows automatically retextured images. The bottom row shows the rigid motion compensated deformation, retargeted to another image.

6.2.5 Paper (ICCV'07) – *Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning*

I46 **Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning**
V. Gay-Bellile, A. Bartoli and P. Sayd
ICCV'07 - IEEE *Int'l Conf. on Computer Vision, Rio de Janeiro, Brazil*, October 2007
Version in French: [N14]
Related papers: [I47,N13]

Most of the image registration algorithms uses a compound cost function which has a data term and a smoother, as described in §5.2.1.1. The data term can be robustified so as to deal with erroneous pixel colors, due to undermodeled phenomena such as external occlusions and shadows. The smoother allows one to coherently fill in the displacement field in those areas where many pixels are discarded. These methods work well

when the expected image displacement warp is smooth. This assumption can be violated in several cases, in particular when the observed surface undergoes a self-occlusion.

The main contribution of this paper is a *pixel-based method for image registration in the presence of self-occlusions*. The idea is to jointly estimate the warp and detect the self-occluded areas. The compound cost function is augmented with a third term we call the 'shrinker'. This forces the warp to shrink on the self-occlusion boundaries. The self-occlusion detection module is based on inspecting the directional partial derivatives of the warp. It is indeed obvious that, at a self-occluded pixel, there exists a direction which makes it vanish. We have shown that this direction can be obtained in closed-form.

An example of this is shown in figure 6.6. Our implementation uses an FFD warp, but the method can be applied to any deformable image warp. The data term is robustified with an M-estimator, so as to deal with both external and self-occlusions.

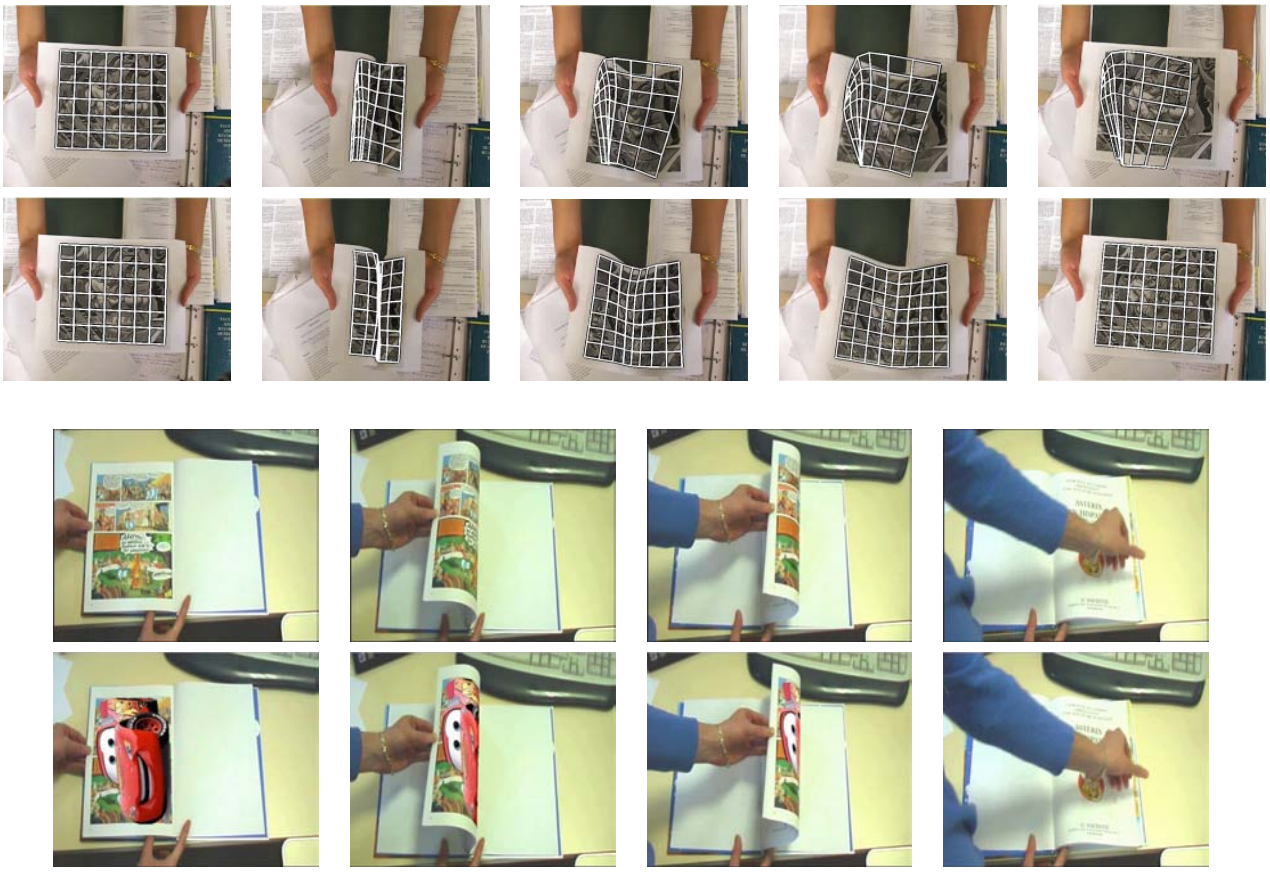


Figure 6.6: Paper (ICCV'07) – *Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning*. The two top rows show a typical self-occlusion example where classical data-robust methods enforcing deformation smoothness fail. The bottom row shows the result obtained with the proposed method. It is seen that the warp collapses onto the self-occlusion boundary as expected. The two bottom rows show an example of retexturing: the "Cars" logo is set up by the user in the reference frame and is then automatically inserted in the other images of the video.

CHAPTER

7

STRUCTURE-FROM-MOTION FOR DEFORMABLE SCENES

In this chapter we study the problem of jointly finding the 3D structure of a deformable environment and the sensor pose from a series of images. We distinguish the single camera and the multiple synchronized camera cases.

In the first part, we assume that a single moving camera observes a moving and deforming environment. The difficult aspect of this is to recover the 3D information. This is because Monocular Deformable SfM is not a naturally well-posed problem. We study various models and priors that combine with the Low-Rank Shape Model so as to find a sensible solution. Our contributions include the estimation of the implicit Low-Rank Shape Model from curve and point correspondences, a robust algorithm that also deals with missing data and priors, and a coarse-to-fine approach.

In the second part, we assume that several synchronized cameras observe the environment. Finding the sparse 3D structure at each time instant is thus almost always a well-posed problem, that can be solved by rigid SfM techniques such as those described in chapter 8. This case can thus be seen as equivalent, to some extent, to having range data. The difficult part is the one of finding the 3D temporal registration. In other words, the general problem of computing the sensor pose and the deformable scene structure is not naturally well-posed. One of our main contributions uses the explicit Low-Rank Shape Model. Another approach we propose discovers the object of interest while registering range images based on quasi-isometric deformations. Finally, we propose a method for the reconstruction of a deformable paper sheet. It is based on a novel parameterization which guarantees that the recovered surface is developable.

7.1 A Single Camera

The works below each comprise various data terms, models and priors. They all are expressed in the framework of multilinear drivers described in §5.3. The most relevant feature which distinguishes these works is which driver they use in particular.

The first three papers propose feature-based methods based on the 2D un-trained driver, also called the implicit Low-Rank Shape Model (LRSM). It is important to recall that, as described in §5.3.2.2, the implicit LRSM does not provide a directly usable 3D structure. If obtaining the 3D deforming structure is the final goal, computing the implicit LRSM should be seen as the leading step in the stratified approach to Low-Rank SfM followed by many authors such as (Bregler et al., 2000) and described in §5.3.4.1. It means that the implicit LRSM is subsequently upgraded to the explicit LRSM, for which specific procedures are proposed in (Brand, 2005; Xiao and Kanade, 2006). The method given in the first paper does not handle missing and erroneous data. A solution to these issues is provided in the second paper. As could be expected, the estimated model generalizes badly when the percentage of missing data is high. The method proposed in the third paper extend the method to incorporate generic prior knowledge based on temporal and spatial smoothness. This dramatically improves the generalization ability of the model.

The fourth paper also uses a feature-based approach but directly estimates the 3D un-trained driver, also called the explicit LRSM. It uses a novel coarse-to-fine approach inspired by the Deformotion concept (Yezzi and Soatto, 2003). The priors proposed in the third paper are used. Finally, the fifth paper uses a 3D pre-trained driver, also called a 3D Morphable Model[1] (3DMM). The fitting is done with a pixel-based data term. The driver is trained with synthetically generated data by the method in (Salzmann et al., 2007b).

7.1.1 Paper (CVPR'04) – *Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences*

I17	**Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences** A. Bartoli, E. von Tunzelmann and A. Zisserman CVPR'04 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Washington, DC, USA, June 2004

Linear features are often used in rigid SfM since they offer a rich source of information, as we report in §8.2. In the deformable environment case, straight lines deform to curves. Our goal is to exploit these features for image registration. The main contribution of this paper is a *method for computing an RBF warp driven by an implicit LRSM based on point and curve correspondences*.

The implicit LRSM estimation process is based on the non-rigid factorization method described as the first step in §5.3.4.1. The problem lies in the fact that while it is pretty easy to obtain curve correspondences, it is much more difficult to match points along the curves. In other words, the curves are parameterized in such a way that corresponding points on the curves have the same parameter.

Our solution proceeds by first applying non-rigid factorization to the point correspondences. This gives the shape bases matrix for these points, and the motion matrix, which holds for any point correspondence in the video. The predicted image points allow us to compute an RBF between the reference image and the other images of the video. We then assess the registration of each curve. A curve is transferred from the reference image to the other images of the video, and the distance to the actual curve is computed. If the average distance is below some threshold, chosen as 0.1 pixels in our experiments, we step forward. If the curve registration is not satisfactory, we introduce a virtual point correspondence on the curve. This point is chosen so as to improve the curve registration. The RBF warps are re-estimated and the process is iterated. Finally, we estimate the shape basis for each virtual point we inserted, and refine all model parameters.

An example of this is shown in figure 7.1. We manually mark points and curves in the first image of the video. Points were tracked with the tracker described in (Shi and Tomasi, 1994). Curves were tracked with a home made pixel-based tracker described in the paper.

[1] Recall that we do not use the appearance counterpart of the 3DMM, but just the shape.

(a) (b) (c) (d) (e) (f)

Figure 7.1: Paper (CVPR'04) – *Augmenting Images of Non-Rigid Scenes Using Point and Curve Correspondences*. The top row shows images from the film "Run Lola Run" retextured with a logo "CVPR 2004" on the forehead of the actress. On this example, there are only a few points that can be reliably tracked around the forehead, but there are several curves such as the hairline and the eyebrows which may be used. The bottom row shows close-up on two images. (a) and (d) show the original images. (b) and (e) show the point and curve correspondences. (c) and (f) shows the automatically retextured images.

7.1.2 Paper (WDV'05) – *A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery*

I22 **A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery**
A. Bartoli and S. Olsen
WDV'05 - *Workshop on Dynamical Vision* at ICCV'05, Beijing, China, October 2005
Other version: [N08]

This paper tackles the implicit non-rigid factorization problem. The main contribution is a *batch method for non-rigid factorization in the presence of missing and erroneous data*. The method is based on the closure constraints that we extended to the non-rigid case. It is thus strongly related to the methods described in §5.5. The method is highly robust since the closure constraints are estimated through RANSAC (Fischler and Bolles, 1981). The eventual reprojection error is minimized to a few pixels, even for highly complex environments. The outliers are generally well detected.

The main drawback of the method is that the recovered model generalizes very badly. In other words, it does not allow one to fill in the data matrix. The main reasons are that the model is empirical and very flexible, and thus tends to overfit the data. The data we processed usually have high ratios of missing data, in the order of 90 to 95%. An example of this is shown in figure 7.2.

7.1.3 Paper (JMIV'08) – *Implicit Non-Rigid Structure-from-Motion with Priors*

J10 **Implicit Non-Rigid Structure-from-Motion with Priors**
S. Olsen and A. Bartoli
Journal of Mathematical Imaging and Vision, special issue: tribute to Peter Johansen, accepted December 2007
Previous version: [I42]

This paper extends the method as described above, where the goal is to improve the generalization ability of the estimated model. This is done by including priors in the estimation. For this, we use temporal and spatial smoothness priors. The main contribution is a *batch algorithm for implicit Low-Rank SfM that handles missing and erroneous data and incorporates priors*. The improvement in generalization is used to fill in the data matrix, *i.e.* to predict the missing data points, and to glue those point tracks which have be split due to imperfect tracking.

Figure 7.2: Paper (WDV'05) – *A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery*. The first two rows show images from a 154 image video from the "Groundhog Day" movie. We automatically tracked 1502 points. The visibility matrix is shown on the third row. It is filled to 29.58%. In other words, more than 70% of the data are missing. The bottom left image shows the points and motion vectors predicted by the model. It shows its extreme flexibility. We used a rank of 15. The final reprojection error is 0.99 pixels and 89.4% of the image points are classified as inliers. The four bottom right images are closeup onto different parts of the scene. They are overlaid with the predicted motion vectors and points (white dots), the original points (light grey squares) and the outliers (dark grey diamonds).

The first prior we propose models temporal smoothness. It uses two assumptions. First, that the camera path is smooth and that its orientation changes smoothly. Second, that the way the environment deforms, which is modeled by the configuration weights, is smooth. Both assumptions hold for most natural videos.[2] Using these together is particularly well suited to the implicit LRSM. The implicit motion matrices indeed depend on both the 'explicit' cameras and the configuration weights. Therefore, a single smoother is used to model both assumptions. It is based on penalizing the finite difference approximation to the first derivative of the implicit camera matrix, through the penalty term:

$$\sum_{i=1}^{n-1} \|\mathsf{J}_i - \mathsf{J}_{i+1}\|_{\mathcal{F}}^2. \tag{7.1}$$

The second prior we propose models spatial smoothness. It is based on the assumption that points consistently close in the images should have close shape bases. This prior is thus particularly efficient for scenes made of a smooth surface. However, experimental results show that it significantly improves the results, even for quite unstructured scenes such as the one shown in figure 7.2. The penalty is written:

$$\sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \alpha(j_1, j_2) \cdot \|\mathbf{K}_{j_1} - \mathbf{K}_{j_2}\|_2^2. \tag{7.2}$$

We sum over the pairs of tracks simultaneously visible in a minimum number of views, say 10. The shape similarity is determined by a Gaussian applied to a distance measure between the tracks chosen as $d(j_1, j_2) = \max_i\{\|\mathbf{q}_{i,j_1} - \mathbf{q}_{i,j_2}\|_2^2\}$.

The method proceeds as follows. An NLS compound cost function must eventually be iteratively minimized. This cost function includes the reprojection error as data term and the two above described smoothers with appropriate weights. In order to find a suitable initialization, we use the implicit Low-Rank SfM method we propose in §7.1.2, and change the implicit coordinate frame so that the temporal smoother is minimized. The reprojection error and the surface shape smoother then give an initialization for the shape bases.

The method improves the generalization error by typically a factor of about 10 in the case of smooth surfaces, and of about 4 in the case of unstructured environments. Preliminary experiments on track gluing show very promising results.

7.1.4 Paper (CVPR'08) – *Coarse-to-Fine Low-Rank Structure-from-Motion*

I50 **Coarse-to-Fine Low-Rank Structure-from-Motion**
A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen and P. Sayd
CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

The main contribution of this paper is a *novel approach to explicit Low-Rank SfM*. This approach differs from the stratified one as it avoids the difficult step of upgrading the implicit to the explicit LRSM. We use a coarse-to-fine ordering of the shape bases in the explicit LRSM. In other words, a shape instance is the combination of a mean shape and gradually finer deformation modes. This was inspired by the Deformotion paper (Yezzi and Soatto, 2003) which defines the motion with respect to both a motion group and an unknown mean shape as the best fit to the data. Deformation is then interpreted as the residuals with respect to this group and possibly noise. This has several computational advantages: the algorithm we propose handles missing data, automatically selects the number of shape bases and makes use of various priors.

The algorithm first uses rigid SfM to recover the camera motion and the mean shape. It then proceeds to incrementally add shape bases. Each step can be efficiently solving by two rounds of LLS and one NLS optimization. The algorithm monitors the generalization ability of the model by computing a v-fold Cross-Validation score, as briefly described in §5.4.1.2. This score typically decreases with the first few added modes. It then starts to increase, and makes the algorithm stop. The algorithm is able to handle the temporal and spatial smoothness priors given by the smoothers (7.1) and (7.2). We take advantage of the mean shape to measure the distance between points. It is perhaps a more sensible measure than the distance between tracks we used in §7.1.3. An example of this is shown in figure 7.3.

[2] Assuming that the video shows a single scene.

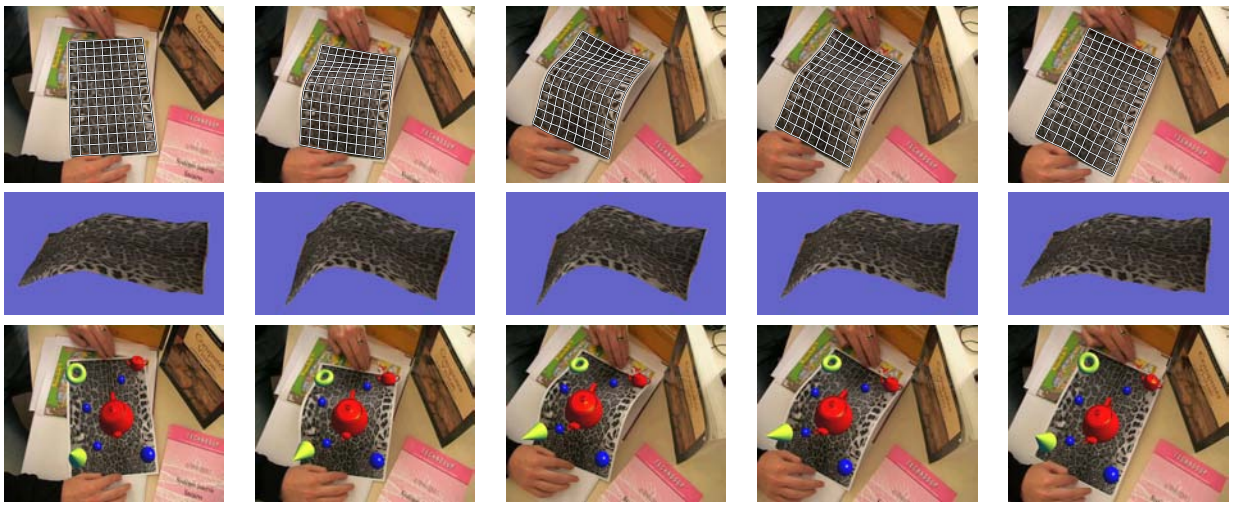


Figure 7.3: Paper (CVPR'08) – *Coarse-to-Fine Low-Rank Structure-from-Motion*. We tracked a paper sheet in a video with the method described in §6.2.4 as illustrated by the visualization grid on the top row. We then applied our coarse-to-fine Low-Rank SfM algorithm which selected 4 shape bases by Cross-Validation and achieved a final reprojection error of 0.84 pixels. The second row shows new views of the reconstructed surface. Finally, the third row shows how virtual objects can be inserted on the reconstructed surface so as to make a full 3D augmentation of the original video.

7.1.5 Paper (ICIP'06) – *Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model*

I29	**Image Registration by Combining Thin-Plate Splines with a 3D Morphable Model** V. Gay-Bellile, M. Perriollat, A. Bartoli and P. Sayd ICIP'06 - *Int'l Conf. on Image Processing*, Atlanta, GA, USA, October 2006

The main originality in this paper is to *combine a TPS warp with a trained multilinear 3D driver, i.e. a 3DMM*. This implements the idea in §5.1 that warps and drivers can be combined together. The former has the advantage of being 'dense' in the sense that it can be applied to any pixel in the image, while the latter allows one to reduce the number of model parameters by incorporating statistical prior knowledge. We use the continuous surface model in (Salzmann et al., 2007b) to generate the training data. We do not model the appearance counterpart.

The minimization strategy follows a forward compositional approach, as described in §5.6. The composition step is achieved by minimizing the difference between transferred and predicted model control points. Local registration is done through a learned linear predictor. The training data are synthetically generated by randomly perturbating the pre-learned driver and the position of the virtual camera.

7.2 Multiple Synchronized Cameras and Range Sensors

In the multiple camera case, recovering the sparse 3D structure at each time instant is generally a well-posed problem, that can be solved by the rigid SfM methods described in chapter 8. The depth is thus obtained for some data points. It is expressed with respect to the local sensor coordinate frame. This section presents four papers which deal with different problems arising when given such range data.

The first two papers use keypoints. The first paper is about the parameterization of paper sheets and its fitting to images obtained by multiple synchronized cameras. The goal is to automatically obtain a physically valid surface that minimizes the reprojection error over keypoints. The second paper uses the explicit LRSM so as to estimate the sensor pose.

The last two papers use dense data, *i.e.* the depth of all image pixels, obtained by stereo sensors. Experiments have also been performed with range data obtained by structured-lighting. The third paper tackles the problem of isometric registration for range data. The method we propose is able to discover the largest isometrically deforming cluster in the scene. The fourth paper registers deformable surfaces based on various smoothers.

7.2.1 Paper (BenCOS'07) – *A Quasi-Minimal Model for Paper-Like Surfaces*

I36 **A Quasi-Minimal Model for Paper-Like Surfaces**
M. Perriollat and A. Bartoli
BenCOS'07 - ISPRS *Int'l Workshop "Towards Benmarking Automated Calibration, Orientation, and Surface Reconstruction from Images"* at CVPR'07, Minneapolis, USA, June 2007
Previous versions: [I27,N07]
Version in French: [N11]

Smoothly bent paper sheets are mathematically modeled by developable surfaces.[3] Their algebraic structure makes these difficult to parameterize. The main contributions we bring are *an intuitive parameterization of paper-like surfaces and an algorithm for its automatic reconstruction*. The long-term research goal is to find technical means for bridging the gap between physical paper sheets and computers, so as to take advantage of the strengths of both worlds. The main assumption needed is that the surface has sufficient texture so as to get enough keypoints. This is weaker than most of the other systems, assuming special lighting conditions and camera pose, such as the one based on shape-from-shading in (Courteille et al., 2007).

The proposed generative model is based on bending a flat surface with appropriate shape along rulings. This is based on the strip approximation to developable surfaces (Pottmann and Wallner, 2001). Our model parameters are ruling positions and bending angles. So as to keep the number of parameters low and naturally generate a smooth surface, we parameterize few rulings only, called guiding rulings. The other rulings, called extra rulings, are interpolated from the guiding rulings. The main advantages of this model are that it can easily be interactively handled by the user and that there is an efficient reconstruction procedure, that we describe below. However, it has a key feature for our practical goals: it easily handles the surface boundaries, which is not the case for many of the other models in the literature.

Our fitting procedure is based on keypoints seen in multiple images. It has three main steps. The first step finds a smooth surface passing close to the keypoints. The second step initializes the model parameters by detecting rulings from the smooth surface. It is known that rulings must not intersect onto the paper sheet. This is used to clean up the detected rulings. The bending angles are estimated from the local surface behaviour. The third step refines all model parameters in a model-based bundle adjustment manner minimizing the reprojection error. This procedure also jointly refines the keypoints while constraining these to lie onto the reconstructed surface. An example of this is shown in figure 7.4.

7.2.2 Paper (ICRA'06) – *Towards 3D Motion Estimation from Deformable Surfaces*

I25 **Towards 3D Motion Estimation from Deformable Surfaces**
A. Bartoli
ICRA'06 - IEEE *Int'l Conf. on Robotics and Automation*, Orlando, Florida, USA, May 2006
Previous version: [I21]

The goal of this work is to compute the pose of a 2.5D sensor with respect to an unknown deformable environment. The method we propose *learns an explicit LRSM and uses it as a 3DMM to compute the pose for new sensor positions*. We directly learn the explicit LRSM. First, the translations are cancelled out by centring the 3D point sets. The rotations are then estimated independently of the deformable structure thanks to the calibrated 3D Low-Rank matching tensors that we propose. These are based on eliminating the structure from the equations. Finally, the configuration weights and shape bases are estimated by factoring some data matrix

[3] Unstretchable ruled surfaces. These surfaces have an everywhere vanishing Gaussian curvature.

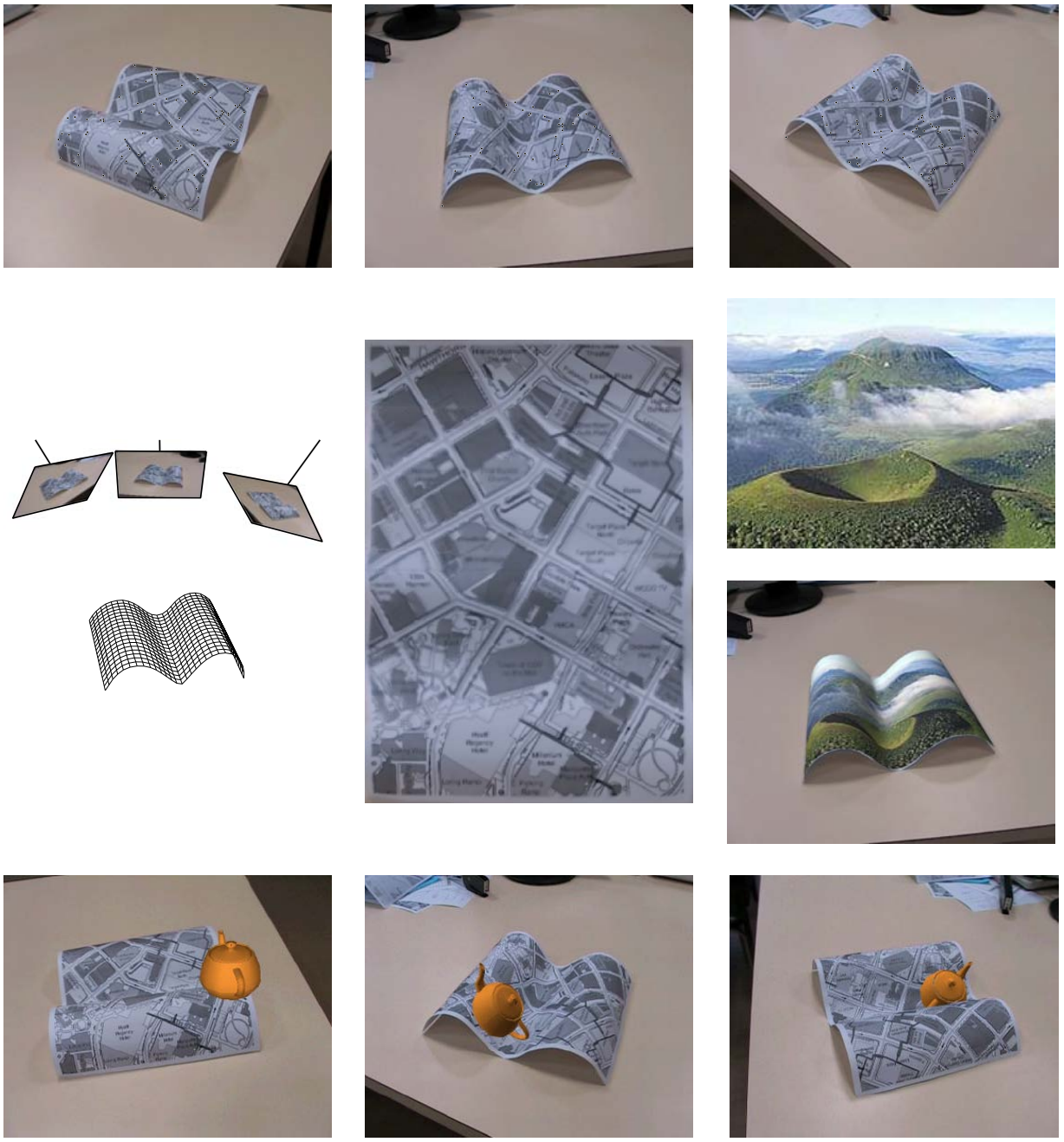


Figure 7.4: Paper (BenCOS'07) – *A Quasi-Minimal Model for Paper-Like Surfaces*. The first row shows three images from a video sequence of a static, bent paper sheet. On the middle row, the left most image shows the reconstructed developable surface, along with the camera pose for the three images of the first row. The middle image shows the texture we recovered by flattening and combining all images from the video (note that the texture is not fully observed in any of the original images). The two right most images show how an image can be used to retexture the surface in the video. The third row shows some frames from a video of a full 3D augmentation of the surface.

using the SVD. Whilst this learning procedure is fast, it requires perfectly matched keypoints. It could easily be extended to deal with missing data by replacing the SVD factorization by one of the methods we proposed in §5.5. This learned explicit LRSM is then use as a 3DMM to compute new sensor pose, and the configuration weights determining the deformation of the environment. Our paper differs from the scene flow computation methods such as the one in (Vedula et al., 2005), or more recently (Pons et al., 2007), which do not recover the (temporal) motion of the sensor, but the displacement of scene voxels.

7.2.3 Paper (BMVA Symposium'08) – *Automatic Quasi-Isometric Surface Recovery and Registration from 4D Range Data*

I48 **Automatic Quasi-Isometric Surface Recovery and Registration from 4D Range Data**
T. Collins, A. Bartoli and R. Fisher
BMVA *Symposium on 3D Video - Analysis, Display and Applications*, London, UK, February 2008

Our goal is to *register dense range images while also discovering the surface of interest and reconstructing a model of the surface*. The main assumption we make is that the surface of interest deforms isometrically. This holds for many cases of interest in practice – quasi-isometry has even been used for the face in (Bronstein et al., 2005). This is motivated by appplications such as object tracking and recognition, texture extraction and remapping and augmented reality. We use dense range images, for which color information is also available. In practice, we get such data with a three camera stereo sensor.

We have to face the problems of data segmentation, registration and aggregation, in the presence of partial data with external and self-occlusions. Our first step uses keypoints. It matches these based on unary constraints such as the distance between SIFT descriptors (Lowe, 2004) and binary constraints. This is inspired by the spectral clustering solution given in (Leordeanu and Hebert, 2005). The binary constraints measure the extent to which two keypoint matches agree with each other. We use a comparison of geodesics measured on the dense range data as binary constraints. The result is an incredibly robust and fast algorithm that matches the keypoints and segments them in isometrically deforming clusters. The second step uses these clusters so as to recover the surface. A deformable surface is 'draped' over the data points. This allows us to flatten each of the patches. The third step recovers the model by compositing the flattened patches together using standard mosaicing techniques. The whole algorithm requires no user interaction and is very fast and reliable. An example of this is shown in figure 7.5.

7.2.4 Paper (3DIM'07) – *Joint Reconstruction and Registration of a Deformable Planar Surface Observed by a 3D Sensor*

I38 **Joint Reconstruction and Registration of a Deformable Planar Surface Observed by a 3D Sensor**
U. Castellani, V. Gay-Bellile and A. Bartoli
3DIM'07 - *Int'l Conf. on 3D Digital Imaging and Modeling*, Montréal, Québec, Canada, August 2007

This paper addresses the joint surface model reconstruction and dense range data registration problems. Contrarily to the above paper, it does not require the color information. There indeed exist sensors that may not give this information such as Time-of-Flight and structured-lighting sensors. The method we propose uses two data terms. The first one is a global attraction of the surface to the data points. The second one attracts the surface boundary to the boundary detected in the data. This term is necessary to prevent the surface to drift away from the data points. These two data terms are robustified thanks to the X84 rule (Castellani et al., 2002). Three penalties are used. The first one encourages spatial smoothness. The second one discourages the surface to stretch. The third one favors temporal smoothness. The estimation algorithm we propose uses Iterated Closest Point (ICP). The closest point computation step is avoided thanks to the use of a distance transform of the range data, as proposed in (Fitzgibbon, 2003). This makes it possible to avoid the traditional two steps in ICP and to minimize the cost function with an efficient NLS algorithm. We use Levenberg-Marquardt. We exploit the high sparsity of the Jacobian and (Gauss-Newton approximation to the) Hessian matrix so as to efficiently solve the augmented normal equations at each iteration. An example of this is shown in figure 7.6.

(a) (b) (c)

Figure 7.5: Paper (BMVA-Symposium'08) – *Automatic Quasi-Isometric Surface Recovery and Registration from 4D Range Data*. (a) shows some inputs range images, (b) shows detected isometric patches that were flattened and (c) compares the true surface texture (left), the rigidly registered patches (middle) and the non-rigidly aligned patches (right).

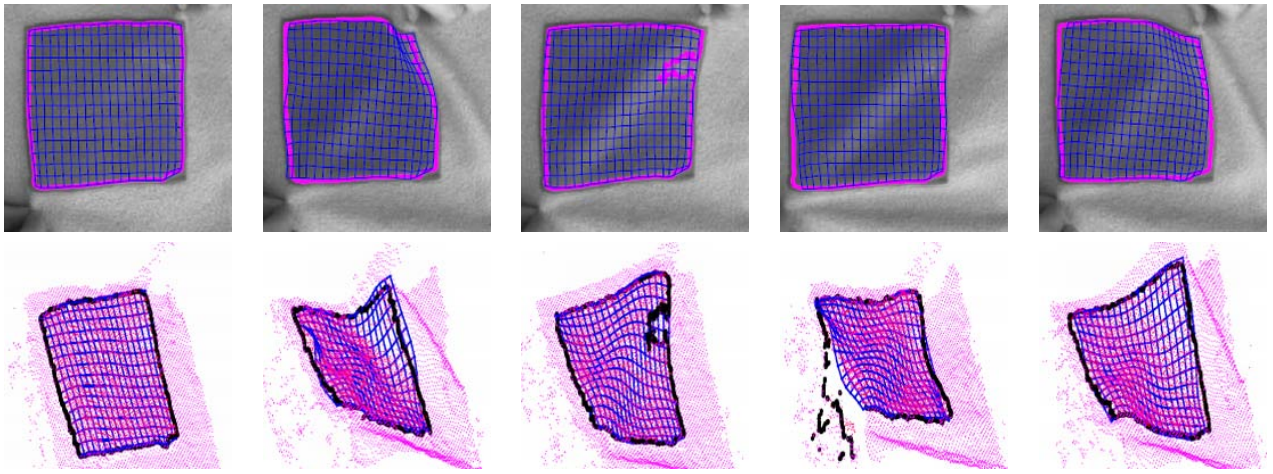


Figure 7.6: Paper (3DIM'07) – *Joint Reconstruction and Registration of a Deformable Planar Surface Observed by a 3D Sensor*. The top row shows images from a stereo video of a deforming cover. The images are overlaid by a visualization grid illustrating the registration. We detect the boundaries as depth discontinuities. As can be seen on the second row, where they are shown as bold curves in the 2.5D data point cloud, they are rather noisy. The registration however manages to reliably register the data and fit to the surface.

CHAPTER

8

STRUCTURE-FROM-MOTION FOR RIGID SCENES

This chapter is about Structure-from-Motion for rigid scenes. The papers are organized in three sections, in terms of the image features they use: points, lines and curves. Various problems are tackled, including camera self-calibration, batch 3D reconstruction and geometry estimation, 3D registration and triangulation. Most of the proposed methods are feature-based. One of our contributions is a composite feature we call Pencil-of-Points, that is based on keypoints on a contour line. A method for pose and instantaneous kinematics from lines observed in a single rolling shutter image is proposed. Finally, we show how 3D curves can be used for quality control in the context of manufactured objects.

8.1 Structure-from-Motion with Points

In general, 'points of interest' represent one of the most versatile, simplest and robust feature one finds in images. Keypoints thus naturally arise in SfM methods. In this section we present three point-based SfM algorithms. The methods use keypoints in their current version, but some of these can easily be modified to deal with other kinds of features. The first paper concerns computing a 3D reconstruction from points correspondences. The proposed method is based on the matrix factorization techniques we have proposed in §5.5. This computes an uncalibrated 3D reconstruction.[1] The second paper concerns camera self-calibration. It more specifically computes the joint constant focal length of a set of cameras given a projective recontruction. This is a nonlinear problem, although our algorithm guarantees to find the optimal solution. The third paper brings a method for computing the affine transformation between two affine 3D reconstructions.

8.1.1 Paper (CVPR'07) – *Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion*

I33 **Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion**
J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert and S. Roy
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

This paper considers as input a set of multiple possibly erroneous image keypoint matches. The proposed method outputs an *uncalibrated 3D reconstruction of points and cameras, and classifies each image point as an inlier or an outlier*. It is well-known that this problem can be formulated as rank-4 matrix factorization with missing and erroneous data. The first factor contains the projection matrices and the second factor contains the structure vectors. They are respectively called the joint projection matrix and the joint structure matrix.

Our algorithm is based on the batch matrix factorization technique we describe in §5.5. It is in fact in this paper that we introduced all the variants of the original approach (Triggs, 1997b). We have to distinguish the affine and perspective camera cases. For both cases, the general factorization technique cannot be directly applied. In the perspective camera case, the data matrix has to be scaled according to the projective depth of the points, which can be recovered from matching tensors such as the fundamental matrix (Sturm and Triggs, 1996). In the affine camera case, the last row of the structure matrix must be made of ones. We call this set of linear constraints the *unity constraints*. This is due to the fact that the problem can be written as rank-3 matrix factorization with an additive affine part. The affine part, corresponding to the translation, can be easily cancelled out for complete data (Tomasi and Kanade, 1992). In the missing data case, it must however be estimated along with the other unknowns. Each of the four methods presented in §5.5 is specifically adapted to deal with the unity constraints:

- ▷ **Method using first-factor closure constraints** (or camera closure constraints). We derive a special form of closure constraints incorporating the unity constraints. The translational part of the cameras is estimated along with their rotational parts.
- ▷ **Method using first-factor basis constraints** (or camera basis constraints). The basis constraints hold on the rotational part of the cameras only. 'Local' translations are estimated for each block while computing the constraints. They are combined together to give the sought after translations as the solution of an LLS problem depending on the computed rotational parts.
- ▷ **Method using second-factor closure constraints** (or structure closure constraints). The unity constraints are enforced while solving for the second-factor from the closure constraints. There are two important modifications to the general algorithm: (*i*) the least singular vector of the design matrix in (5.29) must be discarded and (*ii*) the blocks must not be centred for computing the closure constraints.
- ▷ **Method using second-factor basis constraints** (or structure basis constraints). The bases alignment step is slightly modified but looks similar to the one in the general algorithm since the aligning transformations are projections that take the 'local translations' into account.

[1] A projective 3D reconstruction for perspective cameras and an affine 3D reconstruction for affine cameras.

The resulting algorithms, though slightly different, are almost as simple as the general ones. In can be shown that carefully choosing the coordinate frame in which the 3D reconstruction is expressed is particularly important for the method based on the second-factor basis constraints. Indeed, choosing an orthonormal basis as we propose in §8.1.3, it can be shown that this method minimizes an approximation of the reprojection error.

Extensive experimental results show that combining our methods with iterative NLS miminization is very efficient. The results are accurate since the initial solution provided by our methods almost always allows the iterative ones to reach the global minimum. The computation is efficient since our methods need only two or three rounds of LLS with highly sparse and structured design matrices. Subsequent iterative methods then require few iterations to converge. The methods are robust since each of the constraints is computed using RANSAC. An example of this is shown in figure 8.1.

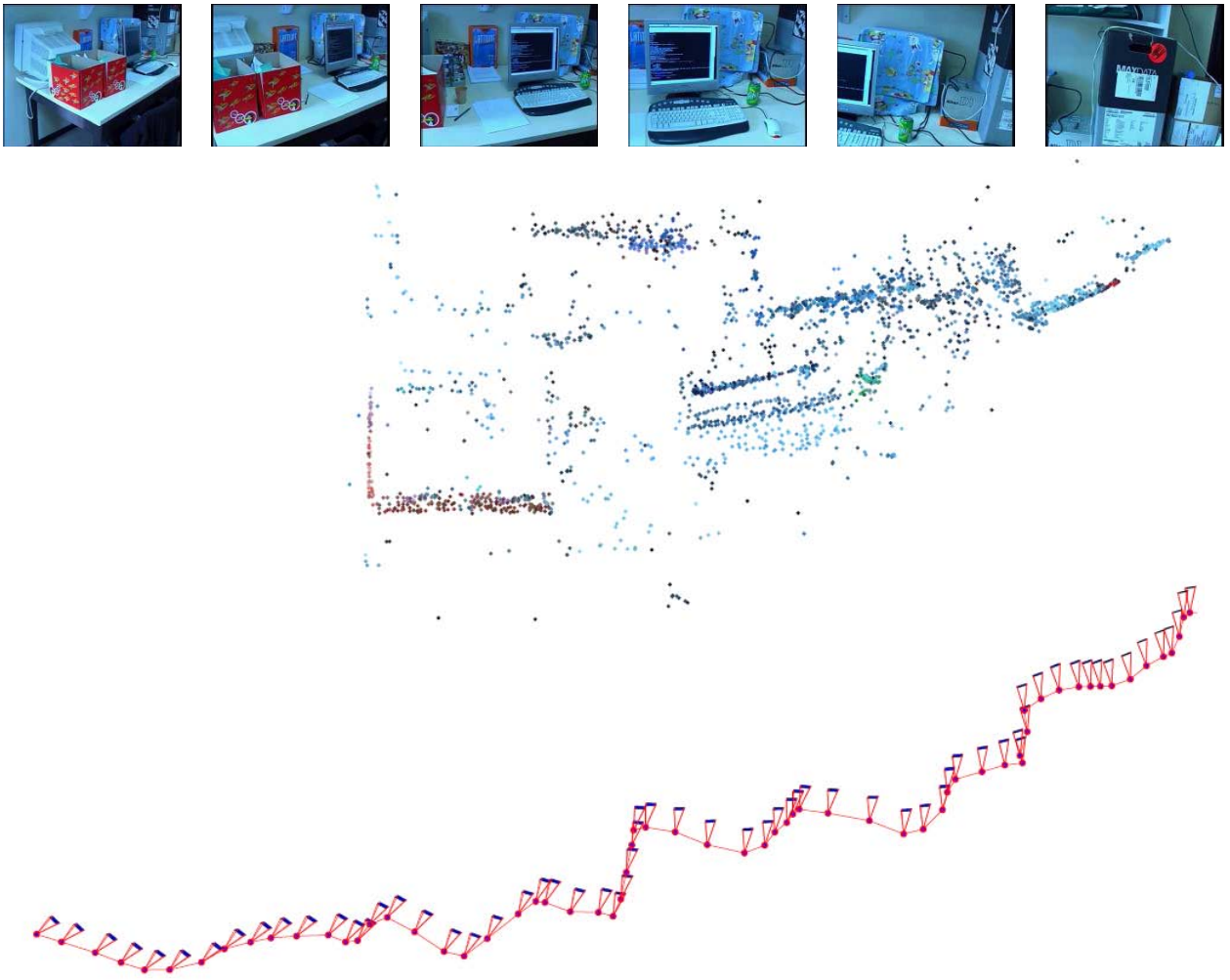


Figure 8.1: Paper (CVPR'07) – *Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion*. The top row shows images out of a 66 image video. The bottom image shows the 3D reconstruction we computed. We used a perspective camera model. The projective reconstruction we obtained was refined by the method in (Mahamud et al., 2001) and was then upgraded to metric using camera self-calibration.

8.1.2 Paper (CVPR'07) – *On Constant Focal Length Self-Calibration From Multiple Views*

I32	**On Constant Focal Length Self-Calibration From Multiple Views** B. Bocquillon, A. Bartoli, P. Gurdjos and A. Crouzil CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007 Version in French: [N10]

We study the camera self-calibration problem in the stratified framework. We assume that the camera focal length is constant and is the only unknown. In practice, we assume that the camera has square pixels and that the principal point lies at the centre of the image. The main contributions of this paper are a *complete study of the Critical Motion Sequences (CMSs)* and an *algorithm that does not have artificial degeneracies*. The problem is tackled within the absolute dual quadric framework (Triggs, 1997a). There are four unknowns: 3 for the plane at infinity in the projective reconstruction and 1 for the absolute conic which directly depends on the constant focal length.

It is important to derive the generic CMSs since they defeat any self-calibration algorithm. A complete classification of the CMSs for the case where all intrinsics are unknown and constant is given in (Sturm, 1997). The case of varying focal length with all other intrinsics known has been extensively studied (Kahl and Triggs, 1999; Pollefeys and van Gool, 2000). Our derivation shows that in the unknown constant focal length case, the focal length cannot be recovered, while the plane at infinity can, if and only if the optical axes of all cameras are parallel. This corresponds to a purely translational camera motion. Note that in the particular case of two cameras with coinciding optical axes, the plane at infinity cannot be recovered either. This also holds for the following other three cases. The first case is when two cameras having coplanar optical axes lie at the same distance to the intersection point. We show that adding a third or a fourth camera as shown in figure 8.2 does not resolve the ambiguity.

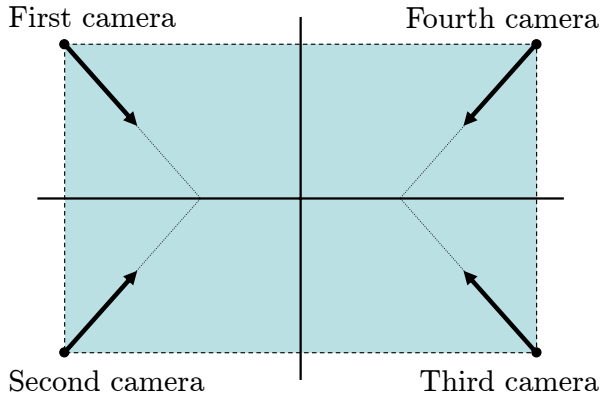


Figure 8.2: Paper (CVPR'07) – *On Constant Focal Length Self-Calibration From Multiple Views*. One of the generic Critical Motion Sequence we have derived for constant focal length camera self-calibration. Using the first two, three or four cameras does not allow one to compute the fixed focal length.

We propose a method for solving the nonlinear self-calibration problem. Its main advantage is that it does not have artificial CMS and does not require an initial solution. Previous methods (Pollefeys et al., 1998; Triggs, 1997a) linearize the problem to find an initial solution and refine it through iterative nonlinear optimization. This introduces artificial CMS, most of which are likely to appear in practice. An example of this is a fixating camera. The algorithm we propose is based on Interval Analysis Global Optimization (Hansen and Walster, 2003). The computational time is in the order of seconds (17 seconds for 4 images in our experiments) and scales linearly with the number of images while being unaffected by the level of noise. We note that Interval Analysis Global Optimization has been used for self-calibration to solve the Kruppa equations in (Fusiello et al., 2004). This approach is however very different from ours. It requires hours of computation and is subject to important singularities due to the Kruppa equations (Sturm, 2000).

8.1.3 Paper (EMMCVPR'05) – *Handling Missing Data in the Computation of 3D Affine Transformations*

I23 **Handling Missing Data in the Computation of 3D Affine Transformations**
H. Martinsson, A. Bartoli, F. Gaspard and J.-M. Lavest
EMMCVPR'05 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, Florida, USA, November 2005
Version in French: [N06]
Previous version: [I20]

In this paper we assume that two sets of cameras sharing keypoints have given rise to two affine 3D reconstructions. It is often the case that they correspond to partial 3D reconstructions of the scene. Finding a complete 3D reconstruction thus involves registering these partial 3D reconstructions. This is a typical step in the hierarchical approach to SfM (Fitzgibbon and Zisserman, 1998). It also is important for one of our batch approaches in §8.1.1. The main contribution of this paper is a *factorization based method that registers two affine 3D reconstructions by minimizing a 'good' approximation to the reprojection error*. The basic method assumes complete data and is extended to deal with missing data using Expectation Maximization (EM) (McLachlan and Krishnan, 1997). The method easily extends to deal with more than two 3D reconstructions.

Our method has three main steps. First, we express each of the two 3D reconstructions in what we call an *orthonormal basis*. This is defined by the joint projection matrix being column orthonormal. The rest of the algorithm would be very similar if this first step were omitted, though the equations would be slightly more complicated. It however makes the cost function that the algorithm minimizes very close to the reprojection error. In the complete data case, the translation can be cancelled out by centring each of the 3D structures. Note that this does not affect the 'orthonormality' of the bases. This is followed by the factorization of a data matrix containing centred image points. In the missing data case, an expectation step is used to predict the missing data points. This is done by reprojecting each of the 3D reconstructions. The rest of the algorithm is similar to the complete data case, except that the image points have to be centred with respect to the reprojected centroid, and not the actual image centroids. Experimental results shows that our algorithm compares favorably with other ones such as the minimization of the transfer error in one set of images only and a direct 3D factorization.

8.2 Structure-from-Motion with Lines

Lines are very commonly found in indoor or outdoor man-made environments. Contour lines have several advantages over keypoints, including that their location is generally more accurate and that their inter-image matching is generally more reliable. Two important drawbacks are that the algebraic representation of 3D lines is non-trivial and that line matches do not allow one to compute the epipolar geometry between two images.

We present three papers. The first one introduces a composite feature called Pencil-of-Points (POP), which is a set of colinear points. We give a complete framework for POP detection, matching and SfM. Note that POPs can be used to estimate the epipolar geometry. The second paper brings an algorithm for triangulating a point lying on a known 3D line. This is a key step in the POP reconstruction framework. Our algorithm finds the global minimum of the reprojection error. The third paper studies the behaviour of scene lines observed by a rolling shutter camera. It shows that knowing the scene model allows one to estimate the relative object to camera pose and kinematics.

8.2.1 Paper (ECCV'04) – *A Framework For Pencil-of-Points Structure-from-Motion*

I16 **A Framework for Pencil-of-Points Structure-From-Motion**
A. Bartoli, M. Coquerelle and P. Sturm
ECCV'04 - *European Conf. on Computer Vision*, Prague, Czech Republic, May 2004

The main contributions of this paper are the *introduction of the POP as a composite image feature and a set of algorithms for POP SfM, including detection, matching, estimation of matching tensors and triangulation.*

POPs have several advantages over keypoints and contour lines including that their matching is more reliable and that they allow one to estimate the epipolar geometry.

POP SfM has the following five main steps, similarly to the two image matching algorithm in (Hartley and Zisserman, 2003). First, POPs are detected in the images. The method we propose is very simple. It uses independently detected contour lines and keypoints. The lines with more than two nearby keypoints are used as POPs. It is observed in practice that the repeatability rate of the POPs detected by this method is higher than that of the keypoints and contour lines it uses as inputs. Second, the detected POPs are matched between the images. We first hypothesis line-level POP matches. For each of these, we try every possible triplets of point matches. Three point matches give a 1D homography relating corresponding points along the supporting line. We use it to compute a cross-correlation score, and select the triplet that maximizes this score. Finally, we use a Winner Takes All scheme to find the final POP matches. The third step is to compute the epipolar geometry. It is seen that only three pairs of POP in general position are required. We use a RANSAC procedure with the 'three POP' algorithm we propose. Fourth, we refine the epipolar geometry or the multiple image geometry if more than two views are used. This is done in a bundle adjustment manner, where the points move along the supporting lines than are also tuned, for all the inlying POPs. Finally, we do a guided matching step. This boils down to using the algorithm in (Schmid and Zisserman, 1997).

An example of this is shown in figure 8.3. We manually calculated the repeatability rate on this example. We obtained 51% for the POPs, 41% for the keypoints and 37% for the contour lines.

Figure 8.3: Paper (ECCV'04) – *A Framework For Pencil-of-Points Structure-from-Motion*. The images go by pairs. The top left image pair shows the detected POPs. The top right image pair shows the 9 putative matches we obtain. Here, there is no matching error. The bottom left image pair shows epipolar lines illustrating the estimated epipolar geometry. The bottom right image pair shows the 11 epipolar geometry guided matches we finally obtain.

8.2.2 Paper (IVC'08) – *Triangulation for Points on Lines*

J09 **Triangulation for Points on Lines**
A. Bartoli and J.-T. Lapresté
Image and Vision Computing, Vol. 26, No. 2, p. 315-324, February 2008
Previous version: [I26]

Triangulation is usual done by minimizing the reprojection error, measured as the sum of squared Euclidean distance between predicted and data points. Recent work also considers cost functions based on different norms (see (Kahl and Hartley, 2007) and references therein). Algorithms based on finding the roots of a system of polynomials have been derived for various cases. The general point case is studied in (Hartley and Sturm, 1997) for two images and in (Stewénius et al., 2005) for three images. The point-on-plane case with two images is studied in (Chum et al., 2005). These algorithms find the global minimum of the reprojection error, but do not easily extend to deal with more than a few images. We propose an algorithm that *solves the point-on-line triangulation problem*. It requires computing the roots of a polynomial in one unknown whose degree is a linear function of the number of images. This algorithm works well for a number of views ranging from one to hundreds (we tested with 387 views on a real example). It runs very fast and can thus be embedded in RANSAC. This algorithm is used for POP triangulation in §8.2.1 and for reconstructing the shape bases in our coarse-to-fine low rank SfM algorithm in §7.1.4.

8.2.3 Paper (CVPR'07) – *Kinematics From Lines in a Single Rolling Shutter Image*

I35 **Kinematics From Lines in a Single Rolling Shutter Image**
O. Ait-Aider, A. Bartoli and N. Andreff
CVPR'07 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007

CMOS cameras are usually low-cost, use low-power and can achieve high frame rates. They often shoot in a rolling shutter mode: instead of acquiring the whole image at once as standard cameras do, they acquire each scanline at once. This makes moving objects to appear distorted in the image. A recent work (Ait-Aider et al., 2006) shows that these distortions can be used to compute the pose and kinematics (*i.e.* an instantaneous 3D rotation and translation), given the scene model as a set of 3D points.

Our paper shows that the same is possible with lines, and proposes an algorithm to *compute the pose and kinematics from lines seen in one rolling shutter image*. The rolling shutter makes 3D lines appear as curves in the image. Our algorithm takes image curves to scene model lines matches as inputs. We define an image curve to be a set of contour points. For each of these points, we instantiate a point lying on the corresponding model line. We then minimize the reprojection error over these contour points. The projection model is a function of the scanline exposure time, that we compute as a function of the image height and camera frame rate. The unknowns are the pose, the kinematics and the shift of each point along its supporting model line. The nonlinear minimization is performed efficiently by using the sparse block structure of the normal equations, which is very similar to the structure found in bundle adjustment. The pose parameters are initialized such that the object lies in front of the camera. The kinematics parameters are randomly initialized with zero speed. The point shift parameters are initialized to zero. An example of this is shown in figure 8.4. The images have size 640×480. They were acquired at 30 frames per second. This gives a scanline exposure time of 39.5×10^{-6} seconds.

8.3 Structure-from-Motion with Curves Applied to Quality Control

Curves are a very natural primitive in CAD models, and a rich source of information for many manufactured objects. We present two papers describing methods that compute a 3D reconstruction of curves and match those to a given CAD model, so as to measure the possible defects. One of our motivations is that most of the other optical sensors such as structured light range scanners allow for an accurate reconstruction of surface points but do not precisely locate the object discontinuities. Our system uses a set of images of the object of interest, for which the camera pose and intrinsics are known. In practice they can be either pre-computed using a calibration

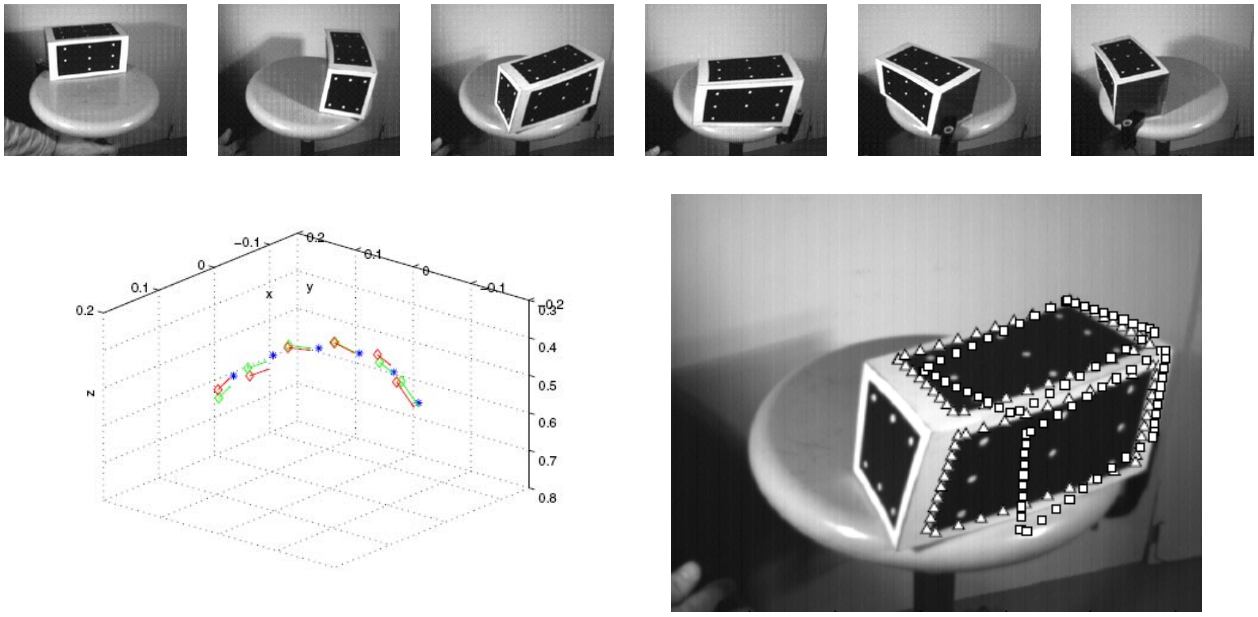


Figure 8.4: Paper (CVPR'07) – *Kinematics From Lines in a Single Rolling Shutter Image*. The top row shows some of the images from the video we took. The distortions are clearly visible given that the real object is a cuboid. The bottom left image shows the pose and kinematics we individually computed for each image of the video. They are consistent with the object motion. The bottom right image shows the reconstructed object model points reprojected onto one of the rolling shutter images with a perspective projection model. This illustrates the correctness of the computed pose, and how an image could be undistorted with this method.

apparatus or computed on-line using uncalibrated SfM and self-calibration as we described in §§8.1.1 and 8.1.2. We roughly align the given CAD model to the images, and refine the curves. We have chosen to use NURBS curves (Piegl and Tiller, 1997) due to their broad use in industrial manufacturing applications. There also are technical reasons such as local control, the possibility of easily inserting control points and the fact that they are projectively covariant. The problem is difficult since the images often exhibit false edges, due for instance to reflections in the case of metallic objects.

There is a significant body of literature on parametric curves in computer vision and photogrammetry. Most of the work on the 3D reconstruction of parametric curves such as (Kahl and August, 2003; Xiao and Li, 2001) uses 'linear' splines such as the cubic B-spline (5.5).

The two core stages of our iterative algorithm can be stated as follows:

1. **Curve parameter refinement**. Refine the parameters of each curve, keeping its number of control points, hence its complexity, fixed.

2. **Control point insertion**. Insert a new control point and loop back to first 1. This step increases the curve complexity.

There are two key components here: the control point insertion strategy and the stopping criterion. We have tried several strategies for the former. The one we eventually chose is to place the point in the knot inverval that has the highest median fitting error. This can be seen as a robustified version of the method in (Dierckx, 1993). The stopping criterion is important since we must allow the curve to be flexible enough so as it models the defects but without overfitting. Therefore, we somehow have to penalize the complexity of the curve. This can be done using a model selection criterion. We used the Bayesian Information Criterion (BIC) (Schwarz, 1978) that ensures asymptotic consistency. The CAD model is used to identify the parts of the curves which are occluded.

Our two papers mainly differ by the particular data terms used. Our first paper uses a feature-based data term. Image contour points are first searched for, and a geometric distance to the predicted curves is then

minimized. Our second paper uses a pixel-based data term. The data term is related to the image gradient along the predicted curves. The two methods are validated and compared using simulated and real datasets, which reveals that the pixel-based method is in general more accurate.

8.3.1 Paper (SCIA'07) – *Reconstruction of 3D Curves for Quality Control*

I30 **Reconstruction of 3D Curves for Quality Control**
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
SCIA'07 - *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, June 2007
Version in French: [N09]

The specific focus of this paper is the use of a feature-based data term. We describe how both steps in the general iterative scheme are implemented. The curve parameter refinement step includes a search for image contours. We sample the reprojected curves. We then search along the curve normal at each of the sampled points in a range of typically 20 pixels. A score is computed that combines the distance to the sample curve point and the image gradient magnitude and orientation. The point with the highest score is selected. The bounded search range and the distance component in the score make the search robust. The control point insertion step requires us to compute an error for each knot interval. Given that we have found contour points at the previous step, we simply consider their distance to the refined curve. The error for an interval is defined as the median of these distances. The BIC score is used as a termination criterion. It has two parts. One of these penalizes the model complexity and depends on the number of control points. The other one is the fitting error, that we measure as the sum of squared distances from the contour points to the predicted curves.

8.3.2 Paper (EMMCVPR'07) – *Energy-Based Reconstruction of 3D Curves for Quality Control*

I37 **Energy-Based Reconstruction of 3D Curves for Quality Control**
H. Martinsson, F. Gaspard, A. Bartoli and J.-M. Lavest
EMMCVPR'07 - IAPR *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EZhou, Hubei, China, August 2007
Related paper: [I45]

The method presented in this paper uses a pixel-based data term; namely the sum of the image gradient magnitude along the reprojected curves. It takes the same steps as in the feature-based version. The difference shows up at the curve refinement step. Instead of minimizing the distance to the detected contour points, we minimize some function of the image gradient. Roughly speaking, we minimize the magnitude of the image gradient vector projected onto the normal vector to the curve, so as to incorporate spatial continuity information. An example of this is shown in figure 8.5.

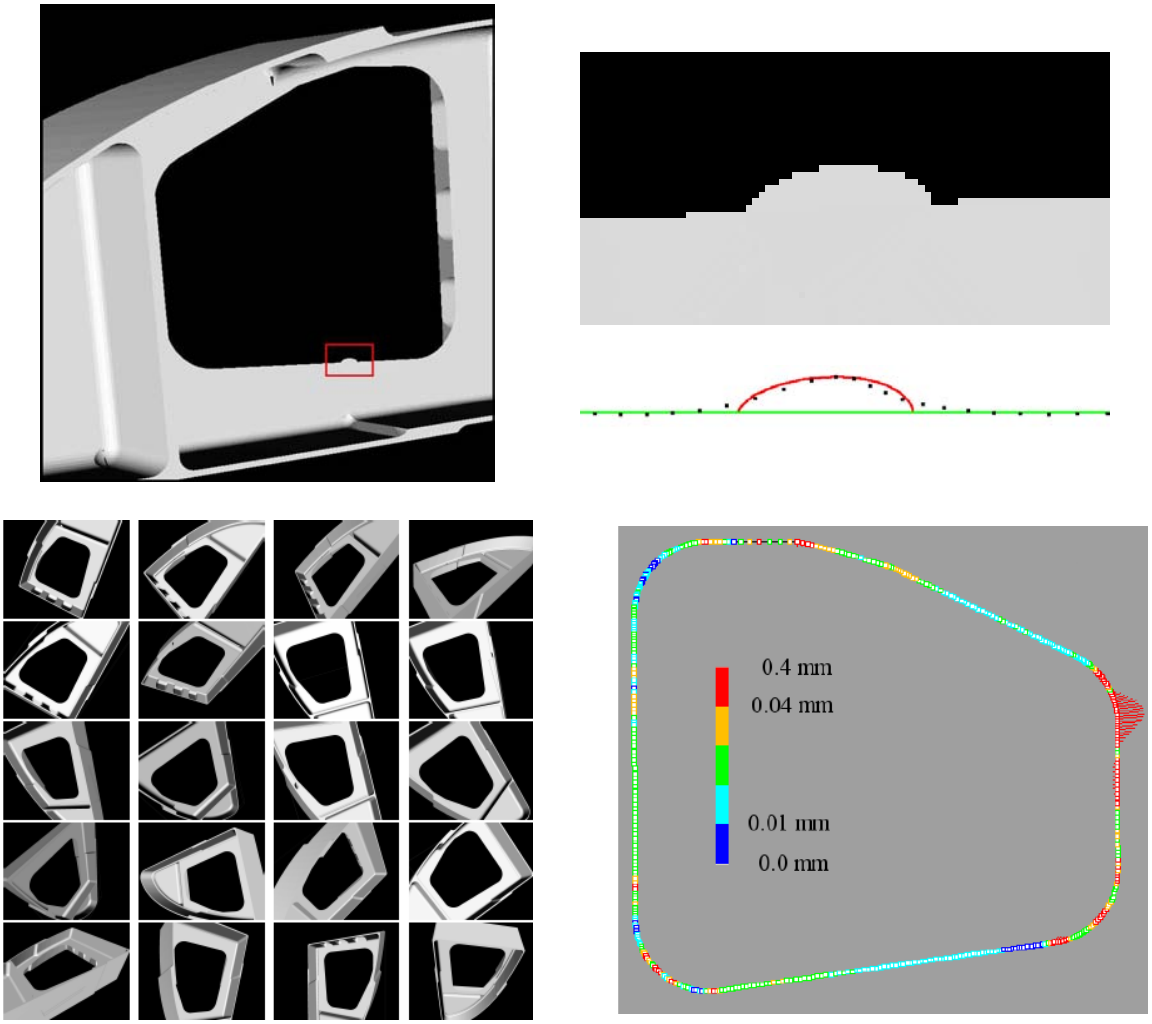


Figure 8.5: Paper (EMMCVPR'07) – *Energy-Based Reconstruction of 3D Curves for Quality Control*. The top left image shows an image synthesized with a CAD model of a car door. The top right images respectively show a closeup on a simulated defect, the original CAD model curve in light grey (green) and the reconstructed curves in dark grey (red). The bottom left thumbnails are some of the 36 real images of the manufactured object. The bottom right image shows the discrepancy between points onto the original CAD model curves and the reconstructed curves. The discrepancy is represented by lines with length proportional to the curve-to-curve distance. A scale factor of 20 is used to ease visualization.

CHAPTER

9

OTHER WORKS

This chapter contains the work I and other collaborators have done that is relevant to several other parts of this document, but which does not naturally fit into a distinct section. In this chapter, two main topics are focused on.

The first one is the Active Appearance Models. We study the difficult problem of person-independent Active Appearance Model fitting. Experimental results reveal that, although efficient in the person-specific context, standard Active Appearance Models are not well adapted to the person-independent context. We propose a solution that we call the multi-level segmented Active Appearance Model. Our second contribution concerns the lighting problem. Explicitly learning the set of possible face appearances under different lighting and (externally) cast shadows is not feasible. Our solution is based on fitting the Active Appearance Model in a light invariant space instead of the regular color space.

The second topic we address is concerned with the Prediction Sum of Squares statistic and Leave-One Out Cross-Validation. We derive non iterative formulaes for non standard Linear Least Squares problems arising in warp estimation. Finally, we show how Leave-One Out Cross-Validation can be used for 2.5D parametric surface reconstruction.

9.1 Active Appearance Models

An Active Appearance Model (AAM) is a 2D statistical model of shape and appearance, originally proposed in (Cootes et al., 2001). The shape counterpart can be seen as an SSM (Statistical Shape Model) described in §5.3.1.2. We are interested in face AAMs. An AAM can be trained to represent variations such as pose, expression, identity and lighting.[1] It is usually fitted to a single image.

We describe one possible way of training an AAM from a set of labelled images. The labels are vertices corresponding to a triangular face mesh model. The first step is to learn the shape subspace. This is done by using PCA (Principal Component Analysis) on the label coordinates, pre-aligned by procruste analysis. A number of leading eigenshapes are kept. Four eigenshapes are added to allow for 2D similarity transformations. The second step is to learn the appearance subspace. This is done by using PCA on the shape normalized images. A number of eigenapperances are kept, and two extra ones are added to allow for gain and bias transformations (5.35). Various face instances can be generated by varying the shape and appearance parameters.

Fitting an AAM consists to find the shape and appearance parameters that make the synthesized image as similar as possible to the input image. This is usually done by iterative NLS optimization. We present two papers. The first one is about person-independent AAM face fitting. In other words, we want to fit an AAM to a face image that has not been included in the training set. The second paper is about AAM face fitting in the presence of lighting change and externally cast shadows.

9.1.1 Paper (BMVC'07) – *Segmented AAMs Improve Person-Independent Face Fitting*

I41 **Segmented AAMs Improve Person-Independent Face Fitting**
J. Peyras, A. Bartoli, H. Mercier and P. Dalle
BMVC'07 - *British Machine Vision Conf.*, Warwick, UK, September 2007

Person-independent face AAM construction and fitting is a challenging problem, mainly due to the high variability in shape and appearance. This paper brings two main contributions. The first one is *to show that the inability of standard AAMs to generate previously unseen faces comes from the appearance counterpart*. The second contribution is what we call *multi-level segmented AAMs*. They are shown to improve over previous AAMs for person-independent face fitting. We consider 40 frontal images of neutral faces all showing a different person.

We start by defining a means to assess fitting accuracy on labelled data. Our Statistical Shape Error (SSE) is based on using several manual labellings of the input image by different users, from which Gaussian statistics are computed. The quality of an AAM fit is assessed using the Mahalanobis distance with manual labelling statistics. The SSE is an essential tool for our experimental analysis.

To determine to best number of eigenshapes and eigenappearances to be kept, so as to maximize the fitting accuracy, we performed the following experiments. We fit several AAMs to our 40 labelled input images and measure the SSE as a function of the number of eigenshapes and eigenappearances. The fitting is initialized by projecting the labels and normalized appearances to the corresponding AAM subspaces. We first tried this in the person-specific context: the input images show a face which is in the training set. Our analysis shows that the best accuracy is reached when all the eigenshapes and eigenappearances are kept. We then tried this in the person-independent context: the input images show a face which is not in the training set.[2] The result is that all the eigenshapes and a number of eigenappearances corresponding to 60% variance of the training data should be kept so as to minimize the loss in accuracy. This is not surprising since in the person-specific, the AAM can fully generate the input image, while in the person-independent one it cannot. Our main statement is that a standard AAM is accurate in the person-specific context but not in the person-independent one. The limitation mainly comes from the appearance component.

Segmented AAMs consist of several partitions, each of which modeling a region of the face such as the mouth. They are more flexible than standard AAMs in that the dependencies between the different segments is weaker than when modeling the face as a whole. We can thus expect that they somehow are able to generalize

[1] Only self-cast shadows are usually learned since externally cast shadows obviously have a too wide variability.
[2] The test is thus done in a leave-one-out manner.

better. The drawback is that fitting each segment independently has a small convergence basin compared to a standard AAM. The multi-level segmented AAM we have proposed is a set of coarse-to-fine segmented AAMs. The fitting strategy gradually splits a standard AAM into these pre-defined segments. We show that this strategy outperforms standard AAMs in the person-independent context and the refitting solution from (Gross et al., 2006). In practice, our multi-level segmented AAM has three levels. An example of the segmented AAM is shown in figure 9.1

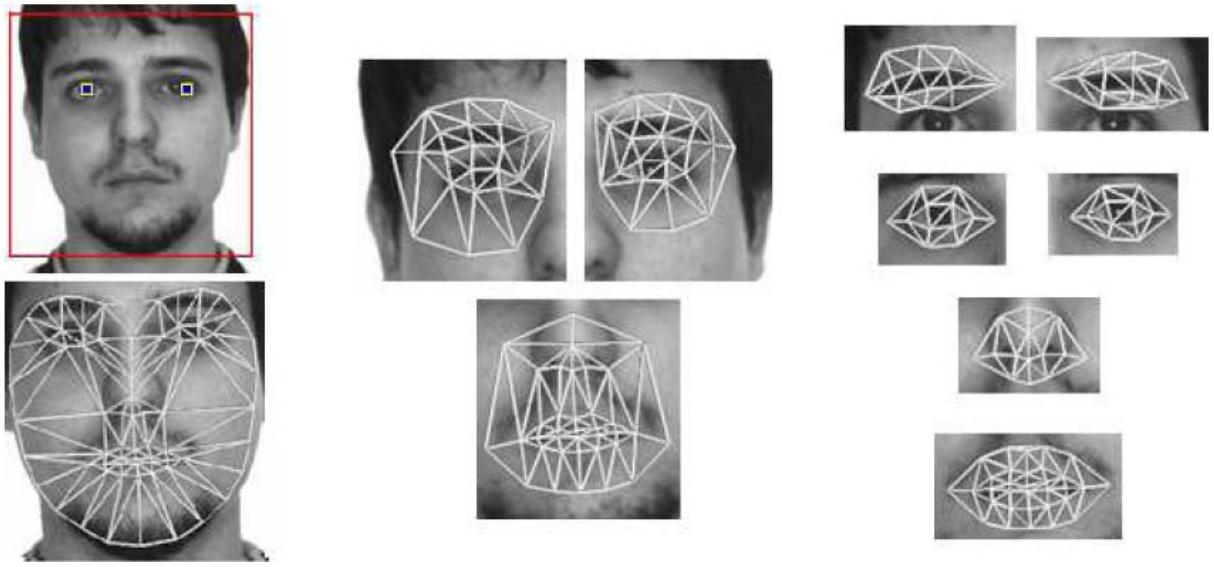


Figure 9.1: Paper (BMVC'07) – *Segmented AAMs Improve Person-Independent Face Fitting*. Illustration of the fitting process for our multi-level segmented AAM. The top left image shows the eye detected with the method in (Fasel et al., 2005) and used to initialized the coarsest AAM shown in the bottom left image. The three intermediary AAMs are then launched, as the middle images illustrate. Finally, the five finest, local AAMs are fitted to accurately fit specific facial features.

9.1.2 Paper (CVPR'08) – *Light-Invariant Fitting of Active Appearance Models*

I49	**Light-Invariant Fitting of Active Appearance Models** D. Pizarro, J. Peyras and A. Bartoli CVPR'08 - IEEE *Int'l Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008

There has been several recent attempts at extending the standard face AAM to deal with lighting variations. Indeed, moving the lighting sources makes a face to self-cast shadows, mostly due to the nose, thereby changing the appearance of the face. In practice, an AAM which has not been specifically tuned to handle lighting variations does not find the expected solution. Existing solutions range from modeling global illumination (Matthews and Baker, 2004), robustifying the cost function to get rid of the illumination induced changes such as shadows (Yan et al., 2003), using a training dataset including a large amount of lighting variations (Sim and Kanade, 2001) and considering independent identity and lighting appearance bases (Kahraman et al., 2007). These solutions are not fully satisfying in unconstrained lighting contexts since they can not model externally cast shadows.

We take a different direction. It is based on the light invariance theory of (Finlayson et al., 2002), that we used in §6.1.3 for image registration purposes. The idea is to *avoid an explicit modeling of the appearance change due to lighting variations*. Recall that color images can be projected to a 1D light invariant space. This transformation has one scalar parameter θ that must be computed, which can be seen as a photometric calibration camera parameter. Light invariant fitting is achieved by comparing the AAM synthesized image and the input image in the light invariant space. This requires one to estimate θ_1 and θ_2. In other words to

photometrically calibrate the camera used to take the training images, and the camera used to take the input image. By slightly abusing the expression, we call these two tasks 'photometric calibration of an AAM' and 'photometric calibration of the input image'. We propose the following solutions:

- ▷ **Light invariant AAM fitting**. We assume that θ_1 is known, *i.e.* the AAM is photometrically calibrated (see below). In practice, θ_2 is rarely known. We estimate it while fitting the AAM in the light invariant space.
- ▷ **Photometric 'self-calibration' of an AAM**. The AAM can be 'self-calibrated' by fitting an input image. If the training and the input images where taken by the same camera, $\theta_1 = \theta_2$ can be estimated while fitting the AAM as above. If two different cameras were used, we generally have $\theta_1 \neq \theta_2$, and have to estimate both while fitting the AAM. It is strictly necessary in both cases that changes in illumination are present between the training and the input images.

Examples are shown in figure 9.2.

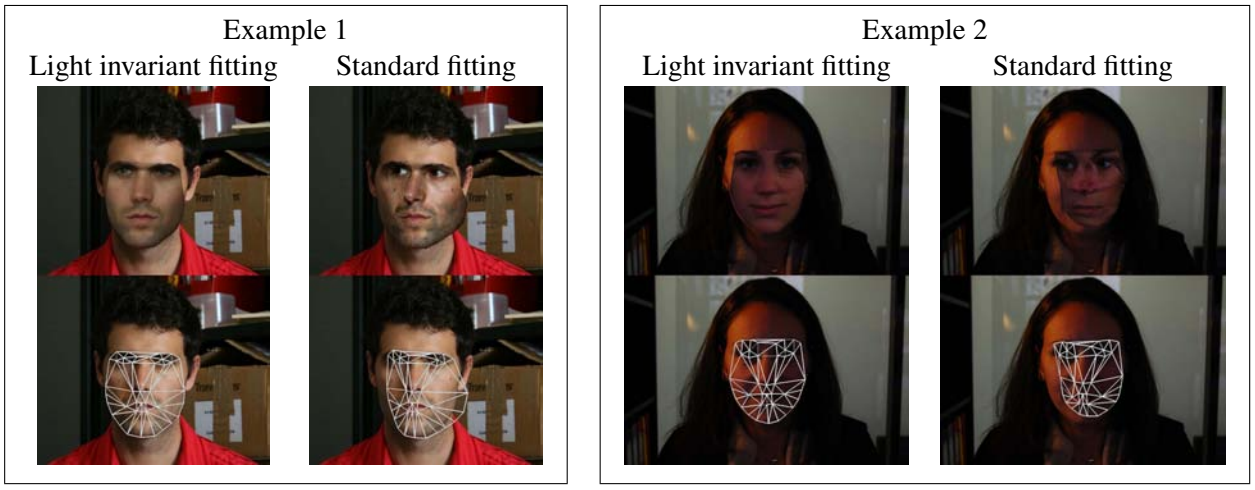


Figure 9.2: Paper (CVPR'08) – *Light-Invariant Fitting of Active Appearance Models*. Four examples are shown. For each of these, the left column shows the result obtained with our light invariant AAM fitting procedure, while the right column shows the result of standard fitting. The top row shows the generated face and the bottom row shows the registered mesh.

9.2 The Prediction Sum of Squares Statistic and Cross-Validation

In this section we present two different papers. The first one extends the non-iterative formula for the Prediction Sum of Squares (PRESS) Statistic and Leave-One-Out Cross-Validation (LOOCV) to non-standard LLS problems arising in some image warp estimation problems. The second paper shows how Cross-Validation can be used in 2.5D parametric surface reconstruction.

9.2.1 Paper – *On Computing the Prediction Sum of Squares Statistic in Linear Least Squares Problems with Multiple Parameter or Measurement Sets*

On Computing the Prediction Sum of Squares Statistic in Linear Least Squares Problems with Multiple Parameter or Measurement Sets
A. Bartoli
Submitted to IEEE *Transactions on Neural Networks*, December 2007

We derive *extensions of the formulaes (5.25) and (5.26) for some non-standard LLS problems*. This is motivated for using the PRESS and LOOCV for problems such as warp estimation: estimating a 2D warp

as defined by (5.4) indeed is a Multiple Parameter and Measurement (MPM) sets LLS problem. Estimating the Rigid-Affine TPS warp we propose in §6.2.1, and any rigid warp, in an instance of the Multiple Scaled Measurements (MSM) sets LLS problem. For both the MPM and MSM LLS problems, and the Multiple Measurements (MM) sets one, we derive the non-iterative PRESS formula. The proofs are given in the paper. The formulaes for the LOOCV score are obtained as above, by replacing the hat by the influence matrix. They are experimentally validated and illustrated on the estimation of our Generalized Thin-Plate Spline (TPS) warps described in §6.2.1.

Multiple Parameter and Measurement sets. This kind of LLS problems has l sets of parameters and measurements represented in matrix form: L is the parameter matrix and R is the measurement matrix with rows $\mathbf{r}_j$. They both have n columns. Each of them is respectively a parameter and a measurement set, the former linked to the latter through the design matrix A. The cost function is as follows:

$$\mathcal{E}^2_{\text{MPM}}(\mathsf{L}) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left\|\mathbf{a}_j^\mathsf{T}\mathsf{L} - \mathbf{r}_j^\mathsf{T}\right\|_2^2 = \frac{1}{m}\|\mathsf{AL}-\mathsf{R}\|^2_\mathcal{F}.$$

The solution to this problem is:

$$\mathsf{L}^\star_{\text{MPM}} \stackrel{\text{def}}{=} \mathsf{A}^\dagger\mathsf{R}.$$

The PRESS is written:

$$\mathcal{K}^2_{\text{MPM}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left\|\mathbf{a}_j^\mathsf{T}\mathsf{L}^\star_{\text{MPM},(j)} - \mathbf{r}_j^\mathsf{T}\right\|_2^2,$$

and can be computed efficiently with the following non-iterative formula:

$$\mathcal{K}^2_{\text{MPM}} = \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1}-\text{diag}\left(\hat{\mathsf{A}}\right)}\right)\left(\hat{\mathsf{A}}-\mathrm{I}\right)\mathsf{R}\right\|^2_\mathcal{F}, \tag{9.1}$$

which is exactly as (5.25) for the standard LLS case, except that the vector two-norm is replaced by the matrix Frobenius norm. This is demonstrated very easily by following the proof in (Montgomery and Peck, 1992) by replacing the vector by the matrix norm. The intuition is that each column of R is independent, in the sense that the corresponding parameters lie in a different column in L, and that $\|\mathsf{U}\|^2_\mathcal{F} = \|\mathbf{u}_1\|_2^2 + \|\mathbf{u}_2\|_2^2 + \ldots$, where $\mathbf{u}_1$, $\mathbf{u}_2$, ..., are the columns[3] of matrix U. The problem can thus be split into l standard LLS problems, and their PRESS combined together to give (9.1).

We could imagine vectorizing the unknown matrix L and using the standard PRESS formula (5.25). This would underestimate the PRESS since l linked measurements must be held out jointly and not individually.

Multiple Measurement sets. We investigate the case where there is a single parameter vector with multiple measurement sets. In other words, each model prediction matches several measurements. This is modeled by the following cost function:

$$\mathcal{E}^2_{\text{MM}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\|\mathsf{CL}-\mathsf{R}\|^2_\mathcal{F} \quad \text{with} \quad \mathsf{L} = \mathbf{u}\mathbf{1}^\mathsf{T},$$

where C is the m row design matrix. This is a particular case of the multiple scaled measurement sets described below. We can obviously not apply the standard PRESS formula for the same reason as above.

The solution is obtained from (9.4) with $\boldsymbol{\omega} = \mathbf{1}$ as:

$$\mathbf{u}^\star_{\text{MM}} \stackrel{\text{def}}{=} \mathsf{C}^\dagger\mathsf{R}\mathbf{1}.$$

[3] This obviously also holds with the rows.

The PRESS is defined by:

$$\mathcal{K}^2_{\text{MM}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left\|\mathbf{c}_j^{\mathsf{T}}\mathbf{u}^{\star}_{\text{MM},(j)}\mathbf{1}^{\mathsf{T}} - \mathbf{r}_j^{\mathsf{T}}\right\|^2,$$

with $\mathbf{c}_j$ and $\mathbf{r}_j$ the rows of C and R respectively. The non-iterative PRESS formula we derive is:

$$\mathcal{K}^2_{\text{MM}} = \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1}-\text{diag}\left(\hat{\mathsf{C}}\right)}\right)\left(\hat{\mathsf{C}}\mathsf{R}\mathbf{1} - \mathsf{R} + \text{diag}\left(\text{diag}\left(\hat{\mathsf{C}}\right)\right)\mathsf{R}\left(\mathrm{I}-1\right)\right)\right\|^2_{\mathcal{F}}. \tag{9.2}$$

Specializing equation (9.6) with $\boldsymbol{\omega} = \mathbf{1}$ to get (9.2) is straightforward.

Multiple Scaled Measurement sets. This case generalizes the previous one by incorporating a different scale for each of the measurement sets, *i.e.* for each column in R, through an $(n \times 1)$ scaling vector $\boldsymbol{\omega}$:

$$\mathcal{E}^2_{\text{MSM}}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{m}\|\mathsf{C}\mathsf{L} - \mathsf{R}\|^2_{\mathcal{F}} \quad \text{with} \quad \mathsf{L} = \mathbf{u}\boldsymbol{\omega}^{\mathsf{T}}. \tag{9.3}$$

The solution is:

$$\mathbf{x}^{\star}_{\text{MSM}} \stackrel{\text{def}}{=} \mathsf{C}^{\dagger}\mathsf{R}\boldsymbol{\omega}. \tag{9.4}$$

The PRESS is defined by:

$$\mathcal{K}^2_{\text{MSM}} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{m}\left\|\mathbf{c}_j^{\mathsf{T}}\mathbf{u}^{\star}_{\text{MSM},(j)}\boldsymbol{\omega}^{\mathsf{T}} - \mathbf{r}_j^{\mathsf{T}}\right\|^2, \tag{9.5}$$

and we propose the following non-iterative PRESS formula:

$$\mathcal{K}^2_{\text{MSM}} = \frac{1}{m}\left\|\text{diag}\left(\frac{1}{\mathbf{1}-\text{diag}\left(\hat{\mathsf{C}}\right)}\right)\left(\hat{\mathsf{C}}\mathsf{R}\boldsymbol{\omega}\boldsymbol{\omega}^{\mathsf{T}} - \mathsf{R} + \text{diag}\left(\text{diag}\left(\hat{\mathsf{C}}\right)\right)\mathsf{R}\left(\mathrm{I}-\boldsymbol{\omega}\boldsymbol{\omega}^{\mathsf{T}}\right)\right)\right\|^2_{\mathcal{F}}. \tag{9.6}$$

This looks like the usual solution (5.25) except that the extra term $\text{diag}\left(\text{diag}\left(\hat{\mathsf{C}}\right)\right)\mathsf{R}\left(\mathrm{I}-\boldsymbol{\omega}\boldsymbol{\omega}^{\mathsf{T}}\right)$ is added to the residual matrix $\hat{\mathsf{C}}\mathsf{R}\boldsymbol{\omega}\boldsymbol{\omega}^{\mathsf{T}} - \mathsf{R}$. This term corrects the bias that would be introduced by using the usual PRESS formula.

9.2.2 Paper (ROADEF'08) – *Reconstruction de surface par validation croisée*

N15	**Reconstruction de surface par validation croisée** F. Brunet, A. Bartoli, R. Malgouyres et N. Navab ROADEF'08 - *Journées de recherche opérationnelle et d'aide à la décision*, Clermont-Ferrand, France, février 2008

In this paper we consider the problem of fitting a 2.5D surface to a sparse set of 2.5D points. We show that *the Cross-Validation score allows finding a sensible trade-off between goodness of fit and surface smoothness*. The surface model is an $\mathbb{R}^2 \mapsto \mathbb{R}$ parametric function. This is very similar to the $\mathbb{R}^2 \mapsto \mathbb{R}^2$ image warps described in §5.2.2. We use the projection of nonlinearly lifted coordinates as a surface model, as equation (5.4) describes for image warps. Recall that this includes models such as Radial Basis Functions (RBF) and Free-Form Deformation (FFD) like, tensor-product spline surfaces. Our procedure runs fast thanks to an implementation that takes advantage of the sparsity of the design matrices. It selects sensible smoothing parameters.

CHAPTER

10

CONCLUSION AND FUTURE WORK

In the first part of my thesis I have provided a review of my related research and administrational activities. In this second part of my thesis, I have given an overview of my research results over the period 2004 – 2007. The cited papers are collected in a companion document to complete this thesis. My conclusions and perspectives regarding my research activity are now given in this chapter. They match those given in §§3.1 and 3.3 in the first part of this thesis. Other conclusions oriented toward how this all fits in the research structures at different scales (internally to LASMEA, locally in Clermont-Ferrand and at the national and international levels) are given in §3.2, in the first part of this thesis.

My research contributions and results concern the matching of images and the 3D reconstruction of structure and motion for rigid and deformable scenes. My perspectives in these areas on the short and middle terms are detailed in §§10.2 and 10.3, along with how they link to the Post Doctoral fellow and PhD students I am currently supervising or co-supervising. Most of these perspectives are related to the general problem of automatically tuning the complexity of a model, described in §10.4. This is a topic that I foresee as a guiding one for my future work in §10.1.

Three of the PhD students I am cosupervising have started writing up their thesis. We are planning that they defend around september 2008. Mathieu Perriollat has worked on the modeling, parameterizating and 3D reconstruction of paper-like surfaces. We are thinking of using the proposed methods as a starting point for making a prototype, toward transferring the technology to a company, as §10.5 reports. Vincent Gay-Bellile has mostly worked in deformable image registration. Perspectives on this topic are given below. Finally, Hanna Martinsson has been working on rigid SfM for quality control. We obtained promising results based on parametric curve 3D reconstruction. I have also supervised visiting PhD students such as Jean-Philippe Tardif who was with the University of Montréal. Perspectives on our work are given in §10.6.

10.1 Transversal Topics

In consolidating all my research efforts into a single thesis, it has become clear that a topic is actually shared by most of my recent work: the automatic selection of model complexity. This is a difficult general problem that occurs in several image registration and SfM problems in rigid and deformable environments. I now want to explicitly address this problem in the computer vision framework, and to generalize the proposed methods and results. This new topic is stated below. The two first ones, namely *image matching* and *three-dimensional reconstruction*, are kept the same as written in chapter 4:

3. **Model complexity tuning.** Given a 'flexible' model (for instance a model for which the number of parameters can be changed and where some of the parameters can be constrained by smoothers), how to find the 'optimal' model complexity (*i.e.* the optimal number of parameters and the optimal smoothing weights)?

This new topic is particularly important for vision in deformable environments. It is also related to problems in rigid environments. Numerous details are given below.

More generally, there are several Machine Learning tools that are useful to our problems, such as dimensionality reduction, kernel PCA, manifold learning and feature selection, that I also want to thoroughly investigate.

10.2 Rigid Structure-from-Motion with Camera Self-Calibration and Prior Knowledge

Computer vision is now used in embedded systems as a complement to a Global Positioning System (GPS). Camera self-calibration is very important since the camera parameters might change. Over time most of the community acknowledges that the main theoretical results in this area have been found. There still however are issues in making systems fully reliable by diagnosing unstable and degenerate situations. One way of improving accuracy and reliability is by incorporating prior knowledge. An obvious type of generic prior for video streams is that the camera trajectory is at least continuous and possibly smooth. The idea is then to combine the traditional reprojection error, which serves as a data term, with a smoother. This has been done before. However, one must take into account that the smoothing parameter, which balances the influence of the smoother, has to be automatically tuned. The smoother should typically be given more weight in unstable configurations. There are two strong research points here:

- ▷ **Real-time tuning of the smoothing parameter.** The smoothing parameter would ideally be adapted to each image in the video stream. Its computation should be fast, so that the whole system can process the images as they come, and for instance give feedback to an autonomous vehicle.

- ▷ **Using multiple smoothers.** As said above, the camera trajectory in video streams is at least continuous but is usually also smooth. In order to really draw on this prior knowledge, it is thus probably necessary to include multiple smoothers with adaptive smoothing weights. This makes the real-time computation even more challenging.

I have started to work on how Cross-Validation can be used for the real-time tuning of the smoothing parameter in sequential SfM with, Michela Farenzena, a Post Doctoral fellow I am co-supervising. These topics are also strongly related to the PhD topic of Julien Michot, whom I have been co-supervising since October 2007.

10.3 Monocular Deformable Image Registration and Structure-from-Motion

Despite significant recent advances, there is still much to be done on Monocular Deformable SfM. The goal is to achieve a comprehensive system which would perform as the efficient rigid SfM ones. The main difference with rigid SfM is that using prior knowledge is mandatory, since it is required to make the problem a well-posed one. I identified the following research points:

- ▷ **Image registration.** Matching images of a deformable environment is generally difficult, due to the lack of geometric constraints such as the epipolar geometry. This is of course strongly related to the 3D shape representation that is being used. For instance, it is commonly admitted that starting with only a sparse set of points allows one to robustly recover the camera pose. The work I have done on the Low-Rank matching tensors makes it possible to improve the point tracking in highly unstructured environments. In the case of a continuous surface such as a paper sheet, however, it is possible to exploit better the photometric constraints, and bridge the gap between recent, highly robust feature-based methods such as those developed at the CVLAB at EPFL, and the accurate pixel-based methods I proposed.

- ▷ **Generic and specific priors.** So as to achieve Monocular Deformable SfM, priors are required. Important priors are what I have called the multilinear statistical drivers, and in particular the un-trained Low-Rank Shape Model (LRSM). I showed that using generic priors such as smoothness of the camera path improves the accuracy and stability of the 3D reconstruction to a large extent. Priors form a continuum: there is no prior being totally generic or specific. I believe that finding novel priors is still an open research, that should be based on imagination.

- ▷ **Sequential processing, model completion and updating.** Deformable models are generally learned from multiple images. An example of this is the one of estimating the shape bases of an LRSM. This has so far only been tackled in batch mode: one assumes that all the images are available from the start. It would however be extremely useful to update a deformable model in a sequential manner as the images come.

Samir Khoualed has started his PhD under my supervision in October 2007. He is working on keypoint based observation functions. I have planned to work on the Statistical Shape Models (SSM). The PhD topic of Dawei Liu, whom I have been co-supervising since September 2007, concerns the use of models from continuum mechanics as priors for deformable surface 3D reconstruction. We are working with the M&M team at LaMI. Finally, I am co-supervising Pauline Julian who is doing an industrial PhD funded by the company FittingBox based in Toulouse. She is working on face tracking and 3D reconstruction with Active Appearance Models (AAMs).

10.4 General Methods for Model Complexity Tuning

As the above mentioned perspectives say, automatic model complexity tuning is a key, critical aspect for many algorithms. There are several possible approaches for automatically tuning the complexity of a model, and much remain to do in this area, including:

- ▷ **Cheap, stable and provable criteria.** There is a need for model predictivity criteria that are cheap to compute, numerically well-behaved and for which there exists a method that guarantees to find a sensible solution. An example of this is Cross-Validation. I have proposed and used non iterative formulaes for estimating deformable image warps. There is however no guarantee that in practice the selected smoothing parameter corresponds to a sensible solution, nor that the Cross-Validation score function has a single minimum. The above mentioned registration and SfM problems demand a computationally cheap criterion and a guaranteed estimation framework. I believe that a promising approach is the one I followed in §9.2.1, that gradually adds control centres to a deformable warp while monitoring the Prediction Sum of Squares (PRESS) statistic.

- ▷ **Combining multiple smoothers.** It is often the case that compound cost functions have one data term and one smoother. A smoother, however, is related to prior knowledge about the problem at hand. Including multiple smoothers is therefore a means to model more complex prior knowledge. This raises the problem of how to efficiently find the smoothing parameters, since not all the existing criteria extend to multiple smoothers (for instance, Cross-Validation does) and the computational expense raises in complexity. Using multiple smoothers might open up possibilities for object recognition, by learning object-class specific smoothers.

- ▷ **The number of free parameters versus the effective number of parameters.** These notions are linked to the two methods of directly adding and removing model parameters and constraining these with smoothers. I have investigated both of these, using respectively Cross-Validation and the PRESS statistic. I believe that they have different properties that should be seen as complementary. Combining them together could proove fruitful.

- ▷ **Integrating model complexity tuning and robustness.** The predictivity criteria I have used so far are not in general robust, in the sense of dealing with erroneous data. While it is not too problematic for some well constrained problems such as rigid SfM, for which RANSAC can be used on the first place to filter out the false feature matches, it still is an open issue in deformable warp estimation problems. An interesting way of research is combining predictivity estimation with robust methods such as RANSAC.

Other possible area of research include the use of predictivity criteria with pixel-based methods. This raises the common 'testing on training data problem' since pixel-based data are very dense and correlated. I am one of the supervisors of Florent Brunet, who started his PhD in October 2007. His topic is in dense 2.5D surface reconstruction. We are working on a predictivity criterion based on the $\mathcal{L}$-curve (Lawson and Hanson, 1974) with the desirable above mentioned properties.

10.5 3D Reconstruction of Paper Sheets

We have studied and contributed to several aspects required to find the 3D structure of a paper sheet from images. These include the mathematical modeling with developable surfaces, their parameterization and estimation algorithms. We believe that a significant body of 3D reconstruction techniques has now become mature. We are foreseeing transferring this technology to viable industral applications. These results were achieved with Mathieu Perriollat, a PhD student who I have been co-supervising.

10.6 Matrix Factorization with Missing and Erroneous Data

The matrix factorization algorithms we initially proposed for rigid SfM were generalized[1] in §5.5. They provide LLS methods to find a solution that usually closely approximates the optimal one. They are computationally efficient and have reached a mature level of development. One of our goals is to write and share a library implementing these algorithms. These results were obtained in collaboration with several researchers, in particular Jean-Philippe Tardif who is now a Post Doctoral fellow at the University of Pennsylvania.

[1] We submitted a paper to a conference on this subject. The first author of which is Jean-Philippe Tardif.

APPENDIX

A

ACRONYMS

The following acronyms for Universities, research institutes and groups and frequently used in this thesis:

ANU	Australian National University
ANR	Agence Nationale de la Recherche French National Research Agency
CAMPAR (at TUM)	Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality Chair for Computer Aided Medical Procedures & Augmented Reality
CNRS	Centre National de la Recherche Scientifique French National Centre for Scientific Research
CVLAB (at EPFL)	Computer Vision Laboratory
DIKU	Datalogisk Institut, Københavns Universitet Department of Computer Science, University of Copenhagen
EPFL	Ecole Polytechnique Fédérale de Lausanne
GRAVIR (at LASMEA)	Groupe Automatique : Vision et Robotique
INRIA	Institut National de Recherche en Informatique et Automatique The French National Institut for Research in Computer Science and Control
IRIT	Institut de Recherche en Informatique de Toulouse
LASMEA	Laboratoire des Sciences de Matériaux pour l'Electronique et d'Automatique
M&M at LaMI	équipe Mécanique et Matériaux du Laboratoire de Mécanique et Ingénieries
TIMS	Technologies de l'Information, de la Mobilité et de la Sûreté
TUM	Technische Universität München
UBP	Université Blaise Pascal, Clermont II
UdA	Université d'Auvergne, Clermont I
VIPS at Uni. Verona	Vision, Image Processing and Sound

Bibliography

H. Aanæs and F. Kahl. Estimation of deformable structure and motion. In *Proceedings of the Vision and Modelling of Dynamic Scenes Workshop*, 2002.

E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In M. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing*, chapter 1. The MIT Press, Cambridge, MA, USA, 1991.

O. Ait-Aider, N. Andreff, J.-M. Lavest, and P. Martinet. Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In *European Conference on Computer Vision*, 2006.

M. Alexa, D. Cohen-Or, and D. Levin. As-rigid-as-possible shape interpolation. In SIGGRAPH, 2000.

D. M. Allen. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16:125–127, 1974.

W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quartely of Applied Mathematics*, 9:17–29, 1951.

S. Avidan. Support vector tracking. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(8): 1064–1072, 2004.

S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, February 2004.

S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, November 2003.

B. Bascle and A. Blake. Separability of pose and expression in facial tracking and animation. In *International Conference on Computer Vision*, 1998.

S. Benhimane and E. Malis. Homography-based 2D visual tracking and servoing. *International Journal of Robotics Research, Special Joint Issue IJCV/IJRR on Robots and Vision*, 26(7):661–676, July 2007.

C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In SIGGRAPH, 1999.

P. T. Boggs, J. R. Donaldson, and R. B. Schnabel. Software for weighted orthogonal distance regression. *acmtms*, 15:348–364, 1989.

F. L. Bookstein. Principal warps: Thin-Plate Splines and the decomposition of deformations. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

T. E. Boult and L. G. Brown. Factorization-based segmentation of motions. In *Proceedings of the* IEEE *Workshop on Motion Understanding*, 1991.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, September 2004.

M. Brand. Morphable 3D models from video. In *International Conference on Computer Vision and Pattern Recognition*, 2001.

M. Brand. A direct method for 3D factorization of nonrigid motion observed in 2D. In *International Conference on Computer Vision and Pattern Recognition*, 2005.

C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *International Conference on Computer Vision and Pattern Recognition*, 2000.

M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *Visualization in Biomedical Imaging*, 1996.

A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.

A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

A. Buchanan and A. W. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *International Conference on Computer Vision and Pattern Recognition*, 2005.

K. Burrage, A. Williams, J. Erhel, and B. Pohl. The implementation of a Generalized Cross Validation algorithm using deflation techniques for linear systems. Technical report, Seminar fur Angewandte Mathematik, July 1994.

U. Castellani, A. Fusiello, and V. Murino. Registration of multiple acoustic range views for underwater scene reconstruction. *Computer Vision and Image Understanding*, 87(3):78–89, July 2002.

G. E. Christensen and J. He. Consistent nonlinear elastic image registration. In *Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001.

H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2):114–141, February 2003.

O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. In *International Conference on Computer Vision and Pattern Recognition*, 2005.

O. Chum, T. Pajdla, and P. Sturm. The geometric error for homographies. *Computer Vision and Image Understanding*, 97(1):86–102, January 2005.

T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham. A trainable method of parametric shape description. In *British Machine Vision Conference*, 1991.

T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *European Conference on Computer Vision*, 1998.

T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

T. F. Cootes, S. Marsland, C. J. Twining, K. Smith, and C. J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *European Conference on Computer Vision*, 2004.

T. Corpetti, E. Mémin, and P. Pérez. Dense estimation of fluid flows. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(3):265–380, March 2002.

J. Costeira and T. Kanade. A multi-body factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, 1998.

F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos. Shape from shading for the digitization of curved documents. *Machine Vision and Applications*, 18(5):301–316, October 2007.

C. de Boor. *A Practical Guide to Splines*. Springer, 2001.

A. Del Bue, X. Lladó, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *International Conference on Computer Vision and Pattern Recognition*, 2006.

P. Dierckx. *Curve and surface fitting with splines*. Oxford University Press, 1993.

G. Donato and S. Belongie. Approximate thin-plate spline mappings. In *European Conference on Computer Vision*, 2002.

J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10:5–12, 1976.

I. Fasel, B. Fortenberry, and J. R. Movellan. Generative framework for real-time object detection and classification. *Computer Vision and Image Understanding*, 98(1):182–210, April 2005.

O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, March 2001.

P. F. Felzenszwalb. Representation and detection of deformable shapes. In *International Conference on Computer Vision and Pattern Recognition*, 2003.

G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *European Conference on Computer Vision*, 2002.

G. D. Finlayson, M. Drew, and C. Lu. Intrinsic images by entropy minimization. In *European Conference on Computer Vision*, 2004.

G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. On the removal of shadows from images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 2006.

M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Computer Vision, Graphics and Image Processing*, 24(6): 381–395, June 1981.

A. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*, 1998.

A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(13-14): 1145–1153, December 2003.

M. Fornefett, K. Rohr, and H. S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, 19(1):87–96, January 2001.

D. Forsyth and J. Ponce. *Computer Vision – A Modern Approach*. Prentice Hall, 2003.

P. Fua. Regularized bundle-adjustment to model heads from image sequences without calibration data. *International Journal of Computer Vision*, 38(2), July 2000.

P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16:35–56, 1995.

A. Fusiello, A. Benedetti, M. Farenzena, and A. Busti. Globally convergent autocalibration using interval analysis. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1633–1638, December 2004.

J. E. Gentle, W. Hardle, and Y. Mori. *Handbook of Computational Statistics*. Springer-Verlag, 2004.

B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging*, 2007.

K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4:133–151, 2001.

G. H. Golub and U. von Matt. Generalized Cross-Validation for large-scale problems. *Journal of Computational and Graphical Statistics*, 6(1):1–34, 1997.

S. Granger and X. Pennec. Multi-scale EM-ICP: A fact and robust approach for surface registration. In *European Conference on Computer Vision*, 2002.

R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(11):1080–1093, 2006.

E. R. Hansen and G. W. Walster. *Global Optimization Using Interval Analysis*. Marcel Dekker, 2003. Second Edition.

R. Hartley and F. Schaffalitzky. PowerFactorization: 3D reconstruction with missing or uncertain data. In *Australian-Japan Advanced Workshop on Computer Vision*, 2003.

R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. Second Edition.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.

D. M. Hawkins and X. Yin. A faster algorithm for ridge regression of reduced rank data. *Computational Statistics & Data Analysis*, 40(2):253–262, August 2002.

H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America A*, 11:3079–3089, 1994.

K. Horn and G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

X. Huang, N. Paragios, and D. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 28(8), August 2006.

S. Ilíc and P. Fua. Using Dirichlet Free Form Deformation to fit deformable models to noisy 3-D data. In *European Conference on Computer Vision*, 2002.

S. Ilíc and P. Fua. Non-linear beam model for tracking large deformations. In *International Conference on Computer Vision*, 2007.

M. Irani. Multi-frame optical flow estimation using subspace constraints. In *International Conference on Computer Vision*, 1999.

D. Jacobs. Linear fitting with missing data for structure-from-motion. *Computer Vision and Image Understanding*, 82:57–81, 2001.

F. Jurie and M. Dhome. Hyperplane approximation for template matching. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(7):996–1000, 2002.

F. Kahl and J. August. Multiview reconstruction of space curves. In *International Conference on Computer Vision*, 2003.

F. Kahl and R. Hartley. Multiple view geometry under the L-infinity norm. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 2007. To appear.

F. Kahl and B. Triggs. Critical motions in euclidean structure from motion. In *International Conference on Computer Vision and Pattern Recognition*, 1999.

F. Kahraman, M. Gokmen, S. Darkner, and R. Larsen. An active illumination and appearance (AIA) model for face alignment. In *International Conference on Computer Vision and Pattern Recognition*, 2007.

K. Kanatani. Geometric information criterion for model selection. *International Journal of Computer Vision*, 26(3):171–189, 1998.

M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988.

S. J. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 2008. To appear.

S. Koterba, S. Baker, and I. Matthews. Multi-view AAM fitting and camera calibration. In *International Conference on Computer Vision*, 2005.

C. L. Lawson and R. J. Hanson. *Solving Least Squares Problem*. Prentice Hall, Englewood Cliffs, 1974.

R. Lehoucq and J. A. Scott. An evaluation of software for computing eigenvalues of sparse nonsymmetric matrices. Preprint MCS-P547-1195, Argonne National Laboratory, 1996.

M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *International Conference on Computer Vision*, 2005.

V. Lepetit and P. Fua. Keypoint recognition using randomized trees. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.

J. Lim and M.-H. Yang. A direct method for non-rigid motion with thin-plate spline. In *International Conference on Computer Vision and Pattern Recognition*, 2005.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

W.-S. Lu, S.-C. Pei, and P.-H. Wang. Weighted low-rank approximation of general complex matrices and its applications in the design of 2-D digital filters. IEEE *Transactions on Circuits and System – I*, 44:650–655, 1997.

B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, 1981.

Q.-T. Luong, P. Fua, and Y. Leclerc. The radiometry of multiple images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(1):19–33, January 2002.

D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

S. Mahamud, M. Herbert, Y. Omori, and J. Ponce. Provably-convergent iterative methods for projective structure and motion. In *International Conference on Computer Vision and Pattern Recognition*, 2001.

S. Marsland, C. J. Twining, and C. J. Taylor. A minimum description length objective function for groupwise non-rigid image registration. *Image and Vision Computing*, 26(3):333–346, March 2008.

D. Martinec and T. Pajdla. 3D reconstruction by fitting low-rank matrices with missing data. In *International Conference on Computer Vision and Pattern Recognition*, 2005a.

D. Martinec and T. Pajdla. 3D reconstruction by gluing pair-wise euclidean reconstructions, or "how to achieve a good reconstruction from bad images". In *International Symposium on 3D Data Processing, Visualization and Transmission*, 2005b.

J. Matas, K. Zimmermann, T. Svoboda, and A. Hilton. Learning efficient linear predictors for motion estimation. In *Indian Conference on Computer Vision, Graphics and Image Processing*, 2006.

I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60 (2):135–164, November 2004.

S. Maybank and P. Sturm. MDL, collineations and the fundamental matrix. In *British Machine Vision Conference*, 1999.

G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., 1997.

D. Metaxas. *Physics based deformable models, applications to computer vision, graphics and medical imaging*. Kluwer International Series in Engineering and Computer Science, 1997.

L. Moccozet and N. Magnenat-Thalman. Dirichlet free-form deformations and their application to hand simulation. In *Computer Animation*, 1997.

D. Montgomery and E. Peck. *Introduction to Linear Regression Analysis*. Wiley, New York, USA, 1992.

J. E. Moody. The effective number of parameters: An analysis of generalisation and regularisation in nonlinear learning systems. In *Neural Information Processing Systems Conference*, 1992.

A. Myronenko, X. Song, and M. A. Carreira-Perpinan. Free-form nonrigid image registration using generalized elastic nets. In *International Conference on Computer Vision and Pattern Recognition*, 2007.

M. Nielsen and P. Johansen. A PDE solution of Brownian warping. In *European Conference on Computer Vision*, 2004.

M. Nielsen, P. Johansen, A. D. Jackson, and B. Lautrup. Brownian warps: A least committed prior for non-rigid registration. In *Medical Image Computing and Computer-Assisted Intervention*, 2002.

O. F. Olsen and M. Nielsen. The generic structure of the optic flow field. *Journal of Mathematical Imaging and Vision*, 24(1):37–53, January 2006.

S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging Vision and Graphics*. Springer-Verlag, 2003.

N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J.Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, 2006.

N. Paragios, Y. Chen, and O. Faugeras, editors. *Handbook of Mathematical Models in Computer Vision*. Springer-Verlag, 2005.

J. Park, D. Mataxas, A. Young, and L. Axel. Deformable models with parameter functions for cardiac motion analysis from tagged MRI data. IEEE *Transactions on Medical Imaging*, 15:278–289, 1996.

A. P. Pentland and S. Sclaroff. Closed-form solutions for physically based modelling and recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(7):715–729, 1991.

L. Piegl and W. Tiller. *THE NURBS Book*. Monographs in Visual Communication. Springer-Verlag, 1997. Second Edition.

J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision*, 76(2):109–122, February 2008.

J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. IEEE *Transactions on Medical Imaging*, 22(8):986–1004, August 2003.

T. Poggio, R. Rifkin, S. Mukherjee, and P. Niyogi. General conditions for predictivity in learning theory. *Nature*, 458:419–422, March 2004.

M. Pollefeys and L. van Gool. Some issues on self-calibration and critical motion sequences. In *Asian Conference on Computer Vision*, 2000.

M. Pollefeys, R. Koch, and L. van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *International Conference on Computer Vision*, 1998.

M. Pollefeys, F. Verbiest, and L. van Gool. Surviving dominant planes in uncalibrated structure and motion recovery. In *European Conference on Computer Vision*, 2002.

J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, April 2007.

H. Pottmann and J. Wallner. *Computational line geometry*. Springer-Verlag, 2001.

M. J. D. Powell and M. A. Sabin. Piecewise quadratic approximation on triangles. ACM *Transactions on Mathematical Software*, 3:316–325, 1977.

I. Reid and D. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.

S. Romdhani and T. Vetter. Efficient, robust and accurate fitting of a 3D morphable model. In *International Conference on Computer Vision*, 2003.

S. Romdhani, A. Psarrou, and S. Gong. Multi-view nonlinear active shape model using kernel PCA. In *British Machine Vision Conference*, 1999.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registation using Free-Form Deformations: Application to breast MR images. IEEE *Transactions on Medical Imaging*, 18(8): 712–721, August 1999.

M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-D tracking. In *International Conference on Computer Vision*, 2007a.

M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3D shape recovery. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1–7, August 2007b.

S. Schaefer, T. McPhail, and J. Warren. Image deformation using moving least squares. In SIGGRAPH, 2006.

C. Schmid and A. Zisserman. Automatic line matching across views. In *International Conference on Computer Vision and Pattern Recognition*, 1997.

I. J. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quartely of Applied Mathematics*, 4:45–112, 1946.

B. Schölkopf, A. Smola, and K. Müller. Non-linear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.

G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In SIGGRAPH, 1986.

J. Shi and C. Tomasi. Good features to track. In *International Conference on Computer Vision and Pattern Recognition*, 1994.

H.-Y. Shum and R. Szeliski. Construction of panoramic mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130, February 2000.

T. Sim and T. Kanade. Combining models and examplars for face recognition: An illuminating example. In *International Conference on Computer Vision and Pattern Recognition*, 2001.

N. Srebro and T. Jaakkola. Linear dependent dimensionality reduction. In *Neural Information Processing Systems Conference*, 2003.

H. Stewénius, F. Schaffalitzky, and D. Nistér. How hard is 3-view triangulation really? In *International Conference on Computer Vision*, 2005.

P. Sturm. A case against Kruppa's equations for camera self-calibration. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1199–1204, October 2000.

P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *International Conference on Computer Vision and Pattern Recognition*, 1997.

P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *European Conference on Computer Vision*, 1996.

R. Szeliski and J. Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3): 199–218, 1997.

T. Tarpey. A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician*, 54(2):116–118, May 2000.

J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12(6):1247–1283, 2000.

D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3D shape and nonrigid motion. *Artificial Intelligence*, 36:91–123, 1988.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

P. H. S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1):27–45, 2002.

L. Torresani, D. Yang, G. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *International Conference on Computer Vision and Pattern Recognition*, 2001.

L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 2007. To appear.

B. Triggs. Autocalibration and the absolute quadric. In *International Conference on Computer Vision and Pattern Recognition*, 1997a.

B. Triggs. Linear projective reconstruction from matching tensors. *Image and Vision Computing*, 15(8):617–625, August 1997b.

B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle ajustment — a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, 2000.

R. Urtasun, P. Glardon, R. Boulic, D. Thalmann, and P. Fua. Style-based motion synthesis. *Computer Graphics Forum*, 23(4):799–812, 2004.

S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 27(3):475–480, March 2005.

R. Vidal and R. Hartley. Motion segmentation with missing data using powerfactorization and GPCA. In *International Conference on Computer Vision and Pattern Recognition*, 2004.

R. Vidal, Y. Ma, and S. Sastry. Generalized Principal Component Analysis (GPCA). IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1–15, December 2005.

G. Wahba. *Splines Models for Observational Data*. SIAM – Society for Industrial and Applied Mathematics, 1990.

G. Wahba and S. Wold. A completely automatic French curve: Fitting spline functions by cross-validation. *Communications in Statistics*, 4:1–17, 1975.

H. Weimer and J. Warren. Subdivision schemes for variational splines. *Approximation theory IX*, pages 345–352, 1998.

Y. Weng, W. Xu, Y. Wu, K. Zhou, and B. Guo. 2D shape deformation using nonlinear least squares optimization. *The Visual Computer: International Journal of Computer Graphics*, 22(9):653–660, September 2006.

R. White, K. Crane, and D. Forsyth. Capturing and animating occluded cloth. In SIGGRAPH, 2007.

O. Williams, A. Blake, and R. Cipolla. Sparse bayesian learning for efficient visual tracking. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1292–1304, August 2005.

J. Xiao and T. Kanade. A linear closed-form solution to non-rigid shape and motion recovery. *International Journal of Computer Vision*, 67(2):233–246, March 2006.

J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. In *International Conference on Computer Vision and Pattern Recognition*, 2004.

Y. J. Xiao and Y. F. Li. Stereo vision based on perspective invariance of NURBS curves. In *International Conference on Mechatronics and Machine Vision in Practice*, 2001.

J. Yan and M. Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *European Conference on Computer Vision*, 2006.

J. Yan and M. Pollefeys. A factorization-based approach for articulated non-rigid shape, motion and kinematic chain recovery from video. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 2008. To appear.

S. Yan, C. Liu, S. Z. Li, H. Zhang, H. Y. Shum, and Q. Cheng. Face alignment using texture-constrained active shape models. *Image and Vision Computing*, 21(1):69–75, 2003.

A. J. Yezzi and S. Soatto. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2):153–167, March 2003.

A. Zandifar, S.-N. Lim, R. Duraiswami, N. A. Gumerov, and L. S. Davis. Multi-level fast multipole method for thin-plane spline evaluation. In *International Conference on Image Processing*, 2004.

L. Zelnik-Manor and M. Irani. Temporal factorization vs spatial factorization. In *European Conference on Computer Vision*, 2004.